12-2017

Towards a Greater Understanding of Post-Translational Modification of Transcription Factor, ArsR, in Helicobacter pylori via Bottom-Up Proteomics: Methods Development for Electrospray and Nanospray Approaches

Theresa Gozzo

Towards a Greater Understanding of Post-Translational Modification of Transcription Factor, ArsR, in *Helicobacter pylori* via Bottom-Up Proteomics: Methods Development for Electrospray and Nanospray Approaches

A thesis submitted in partial fulfillment of the requirement for the degree of Bachelor of Science in Chemistry from The College of William and Mary

By

Theresa Angeline Gozzo

Accepted for _____

_____
Dr. John C. Poutsma, Director

_____
Dr. Mark H. Forsyth

_____
Dr. Christopher J. Abelt

_____
Dr. Rachel E. O'Brien

Williamsburg, VA
December 7, 2017

Table of Contents

Acknowledgments

I would like to thank Dr. John C. Poutsma for the opportunity to contribute to a foundation for bottom-up proteomics research in his lab and for his encouragement in undertaking this project. I have enjoyed the challenge and am so grateful for all that I've learned. Another thank you goes to Dr. Mark Forsyth for the chance to work on such interesting and important research about *H. pylori*. I want to thank Emma Walhout and Matthew Wang for welcoming me to their project and showing me the basics, and I want to especially thank Nita Takanti and Amy Schienschang for their contributions to this work and their perseverance in the frustration of NSI source testing. I would not have accomplished this without their help. I, of course, thank my family and friends for their support throughout this endeavor; they are always behind me, pushing me to reach my goals. Lastly, I would like to thank the Charles Center, the NIH, the NSF, and the donors to my honors fellowship.

List of Figures

List of Tables

Abstract

Studies presented here work towards the analysis of post-translational modification of the essential ArsR transcription factor of *Helicobacter pylori*. A human health hazard, *H. pylori* contributes to the development of gastric cancer and peptic ulcers in its hosts. It is of interest to study the mechanisms that allow the bacterium to survive long term in the harsh stomach environment. The ArsRS signaling pathway mediates the repression of sabA, the gene coding for the adhesin protein, SabA, in acidic conditions. The repression is dependent on histidine kinase, ArsS, but independent of phosphorylation of ArsR at the canonical aspartic acid 52 (D52). To begin comparing the repression mechanism in wild-type and mutant (D52E or D52N) *H. pylori* under neutral and acidic conditions, preliminary bottom-up proteomic investigations were conducted via LC-MS/MS.

Data-dependent mass spectrometric analyses of proteolytic (Asp-N and trypsin) digests of purified ArsR protein were performed with an ESI-LTQ linear ion trap. Ultimately, SEQUEST identified the possible phosphorylation sites, D47, D52, and D59, with high confidence from the CID product ion spectra of the purified ArsR peptides. A shotgun proteomic analysis of *H. pylori* (26695) was also conducted, but it was determined that the combined low-abundance of ArsR and low signal imparted by ESI preclude PTM analysis of similar shotgun samples. In addition, nanospray and electrospray methods were developed for these and future analyses. Drastic improvements in ionization efficiency and signal were observed with the nanospray ionization source in comparison to electrospray. Reliable nano-flow HPLC and nanospray ionization will be essential for future proteomic studies and the continuation of this work.

Chapter 1: Introduction

1.1 Proteomics

Proteomics is the study of the proteome, the entire range of proteins expressed within an organism, tissue, or cell (*1*). In experimental practice, proteomics encompasses both large-scale characterization and targeted protein studies; therefore, a vast array of investigations falls under this heading. As the laborers of the cell, proteins relay signals, repress or activate transcription, catalyze reactions, form structural support, and much more, so it is of great interest to study them. There are three main divisions of proteomics: structural, functional, and expression (*2*). Structural proteomics may include examining any of the levels of a protein's structure or mapping the proteins of different cellular structures (*2, 3*). Functional proteomics aims to elucidate protein function and how that function is carried out (*2, 3).* Goals of determining proteins' interactions with other molecules or identifying post-translational modifications will fall under this category (*2,* 3). Expression, or profiling, proteomics involves quantifying proteins of groups that are subjected to different conditions (*2, 3*). This can mean examining the breadth of expression within a cell or targeting particular cellular components or proteins (*2*).

Though the field of genomics is more mature than that of proteomics, and complete genomes are available for many organisms, studying expressed proteins reveals more about gene products than the genome can alone. As cellular conditions change, transcription is affected; the presence of different regulators alters which genes are transcribed. With that in mind, a single gene can code for multiple mRNAs depending on the exons included in the product (*4*). Translational controls then regulate which and how much protein is synthesized from those mRNAs, and, lastly, post-translational modifications affect final structures and their functioning; thus, the full diversity

of a cell cannot be predicted solely from genetic blueprints. Additionally, acquiring rich proteomic information helps us clarify signaling cascades, recognize misregulation and misfoldings, identify disease biomarkers, and reveal disease pathways. These discoveries can lead to medical advances. Often, proteins are the targets of drug therapies. Furthermore, as the biopharmaceutical field expands, and drug delivery advancements continue to be made, peptides and proteins are increasingly used as pharmaceuticals themselves, providing more reason to explore proteomes (*5-7*).

Mass spectrometry is one of the primary tools used for proteomics. Since proteomics involves the analysis of complex mixtures of peptides and proteins, highly sensitive instrumentation is required. In addition to complexity, many proteins are present at low levels (*8*). The analytical challenge lies in identifying the low abundance proteins in the presence of a variety of more abundant proteins. Sensitive and robust, mass spectrometers have wide working ranges, and, in many cases, are the only instrument that can provide the answers we seek (*9, 10*). Mass spectrometers operate on ions in the gas-phase. The most basic elements of any mass spectrometer are the ionization source, mass analyzer, and detector. The mass analyzer will separate ions based on mass-to-charge (m/z) ratios in some way (differs depending on the analyzer) and the detector will indicate the signal intensity for each m/z value (*11*). Often times the analytes will be subjected to fragmentation, yielding product-ion spectra that can be used to identify them, learn about their structures, and quantify them. We can even monitor chemical reactions by MS.

In terms of proteomics, this means that we can positively identify proteins using high mass accuracy and mass mapping, the fragmentation patterns of the protein or peptides, and database searching (*8*). Not only does this help us determine primary structure, but, once identified, we can gain quantitative data from ion counts. Protein complexes' components can be uncovered in this

way as well, giving us valuable details about protein-protein interactions (*8*). When it comes to post-translational modifications, especially phosphorylation, mass spectrometry is unrivaled in determining site-specific additions (*10*). It surpasses the traditional phosphorylation recognition techniques because it can rapidly assign specific residues' modifications by their gains in mass (*10*). Separation techniques like gel electrophoresis and liquid chromatography can easily be performed prior to MS analysis, decreasing the sample complexity and, thereby, the complexity of the resultant spectra (*9*). Advancements such as ion mobility spectroscopy add another element of separation, allowing proteins with the same primary structure but different folding to be recognized (*9*). Essentially, mass spectrometry remains the standard for proteomics research due to its sensitivity, high mass accuracy, high resolution, and high throughput (*8, 9*).

To analyze proteins via mass spectrometry, one must first isolate them from the biological sample of interest, so the first step in a proteomic experiment is to perform chemical or physical disruption to lyse cells and release their contents (*12*). While mechanical approaches, like beating with glass beads, has been common in the past, it is not as effective in high-throughput and small volume research (*12*). In the end, lysis techniques are tailored to the organism and proteins of study with the downstream analysis methods in mind (*12*). Since some reagents interfere with ionization processes, this is taken into consideration when planning an MS-based analysis (*12*). Once the cell contents have escaped their cell walls and/or membranes, the nucleic acids (DNA and RNA) are broken up by micro-sonication or nuclease addition (*13*).

Finally, the proteins are separated from cell debris, usually through centrifugation. Some proteins exhibit strong physical and chemical interactions with other molecules in the cell that interfere with extraction and isolation methods (*14*). For example, membrane-bound proteins have highly hydrophobic regions which not only need to be separated from the lipids of the membrane,

but need to be solubilized (*14*). It may be even be necessary to separate a particular organelle from the rest of the cell for differential expression or intracellular transport investigations (*14, 15*). One of the biggest obstacles in proteomics is overcoming the protein concentration dynamic range which can be up to nine orders of magnitude in clinical samples (*14*). While performing these preparative steps, one may want to try to equalize concentration by selectively depleting highly abundant constituents or by collecting more balanced ratios of proteins with bead-bound ligand libraries (*14*). To combat this same problem, after extraction and prior to MS introduction, proteins typically undergo some form of separation which reduces the amount or variety of sample components entering the system at a time and thereby improves spectral resolution and ionization efficiency (*3, 11, 16*).

The main separation techniques are gel-electrophoresis and liquid chromatography (*2, 3, 8, 11, 12, 14-16*). Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) is a traditional procedure in proteomics owing to its high-resolution separations of proteins based on molecular mass and charge (*14, 15*). It is unrivaled in separation efficiency by standards of resolution and sensitivity and can separate isoforms resulting from post-translational modifications or alternative splicing (*15*). Intact proteins are made to dissolve and unfold with detergents and chaotropes (H-bond disruptors), like sodium dodecyl sulfate (SDS) or urea, and then are distributed along the gel first by a pH gradient in one dimension and then by an electric potential in the second dimension (*14*). Protein bands are visualized by staining with MS-compatible dyes like Coomassie Brilliant Blue and then analyzed directly by MS or subjected to digestion (*14*).

Once the proteins are isolated and separated from one another, two general mass spectrometric approaches are considered: top-down and bottom-up. So-called “top-down” proteomics involves direct analysis of the protein(s) themselves. This approach requires highly-

specialized mass spectrometers with high mass range and extremely high resolution. Alternatively, one can choose to digest the protein(s) with a proteolytic enzyme and analyze the resulting peptides in the "bottom-up" approach. In contrast to top-down, bottom-up peptide analysis can be performed on a wider variety of mass spectrometers. Below is a diagram specifying the differences in the top-down and bottom-up approaches.

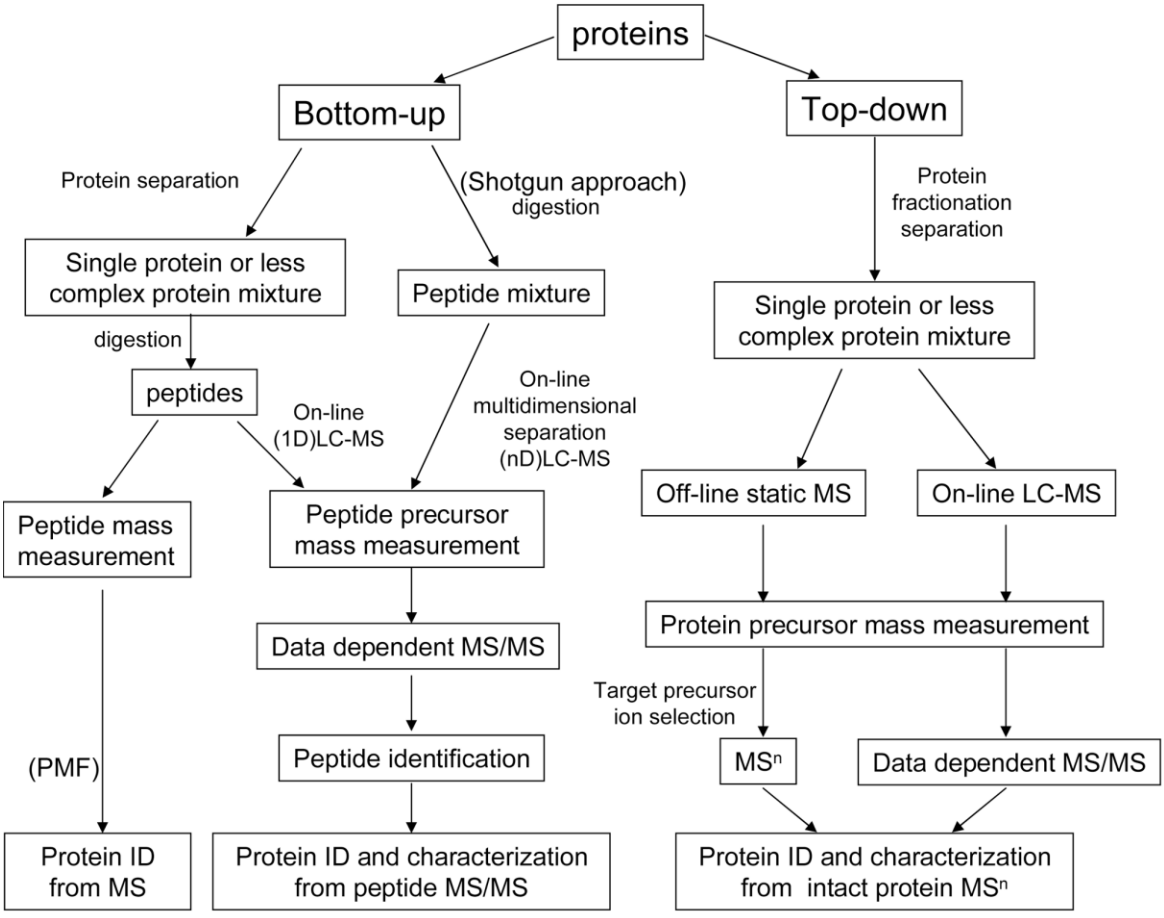

Figure 1.1 Flow Chart Highlighting Some Procedural Differences between Bottom-Up and Top-Down Proteomics (*11*)

1.1.1 Bottom-up Approach

Bottom-up proteomics seeks to identify proteins by the analysis of the peptides created during their proteolytic digestion (*14*). As discussed, after the proteins are acquired from the

biological sample of interest, one can use gel-electrophoresis to separate them to produce less complex mixtures or target a specific protein (*11*). One then digests a sub-set of proteins, or even a single band from the gel. While proteins could be extracted first, digestion is usually performed in-gel. In-gel digestion can be followed by another separation with either 1D or 2D liquid chromatography (LC) (*14*). With the development of LC coupled to tandem MS, 2D-PAGE use has reduced (*14*). LC is more suitable for high-throughput studies due to the labor and time intensity associated with 2D-PAGE (*12, 14,* 15). Additional drawbacks to PAGE include low dynamic range, possible sample loss to the gel, and decreased peptide sensitivity due to SDS contamination (LC/ESI source only) (*14, 15*). 1D-SDS-PAGE, however, has become a common procedure when complemented by LC separation; an increase in the depth of proteome analysis results (*14, 15*). This technique of electrophoretic division, gel slice excision, and in-gel digestion before chromatographic separation is referred to as GeLC-MS/MS (*14, 15*). The benefits can outweigh some of the drawbacks of working with gels.

Depending on the kind of study, one may choose to avoid gels entirely in favor of multidimensional LC or nanoLC. This is the case with shotgun proteomics in which one digests the entire set of proteins immediately after extraction without pre-separation (*14, 16*). LC-MS/MS follows. The shotgun approach is the standard for high-throughput, global profiling investigations, but it has found applications in an extensive range of other protein research (*11, 14, 16*).

To begin digestion, proteins need to be solubilized and unfolded so that the digestion enzyme has more access to the amide bonds for cleavage (*17*). As previously stated, detergents and chaotropes help dissolve and expand proteins, but caution must be exercised so that peptide sensitivity in LC-MS won't be affected (*14*). Available MS-compatible surfactants operate by evaporation or degradation prior to analysis (*14*). In addition, physical methods, such as

microwave heating, aid in digestion efficiency (*14*). Reduction and alkylation are also typically performed. First, dithiothreitol (DTT) is used to reduce disulfide bonds between cysteine resides (*17*). Next, iodoacetamide (IAA) alkylates those thiol groups to prevent disulfide bridges from reforming (*17*). For shotgun studies, reduction and alkylation certainly prove beneficial, but if the investigation targets proteins that possess no cysteine residues, one can avoid this step (*17*). Following reduction and alkylation, proteins are digested into peptides using a proteolytic enzyme. Cleavage occurs by hydrolysis of the amide bond before or after certain residues (*14*). The specificity depends on the particular protease. Trypsin, most commonly chosen for bottom-up protocols, cleaves on the carboxyl side of lysine (K) and arginine (R) residues (*14*), providing two possible sites of protonation for soft ionization techniques (*18*). Alternate proteases, their specificity and common applications can be viewed in Table 1.1, below. At times, improved sequence coverage arises from a multi-protease digestion. Sequence coverage refers to the amount of the primary sequence of the protein represented by peptides in MS spectra.

Table 1.1 Common Proteases Used for Shotgun Proteomics (Adapted From *14*)

Protease	**Cleavage Specificity**	**Common Applications**
trypsin	-K,R-↑-Z- not -K,R-↑-P-	general protein digestion
endoproteinase Lys-C	-K-↑-Z-	trypsin alternative for increased peptide length; multiple protease digestion;^{18}O labeling
chymotrypsin	-W,F,Y-↑-Z- and -L,M,A,D,E-↑-Z- at a slower rate	multiple protease digestion
subtilisin	broad specificity to native and denatured proteins	multiple protease digestion
elastase	-B-↑-Z-	multiple protease digestion
endoproteinase Lys-N	-Z-↑-K-	increased peptide length; create a higher charge state for ETD
endoproteinase Glu-C	-E-↑-Z- and 3000 times slower at -D-↑-Z-	multiple protease digestion; increased peptide length for middle-down proteomics (*19*);^{18}O labeling
endoproteinase Arg-C	-R-↑-Z-	multiple protease digestion
endoproteinase Asp-N	-Z-↑-D- and -Z-↑cysteic acid- but not -Z-↑-C-	multiple protease digestion
proteinase K	-X-↑-Y-	nonspecific digestion of membrane-bound proteins
OmpT	-K,R-↑-K,R-	increase peptide length for middle-down proteomics

[a]B – uncharged, nonaromatic amino acids (i.e., A, V, L, I, G, S); X – aliphatic, aromatic, or hydrophobic amino acids; and Z – any amino acid

An in-gel digestion occurs in the same way as in solution, but the gels require slicing, excising, and de-staining before reduction, alkylation, and proteolysis (*17*). The peptides are then extracted from the gel into solution and prepared for MS or LC-MS analysis (*17*). Following digestion, buffers are removed and samples are resuspended and desalted (salts interfere with ionization) in preparation for MS or LC-MS.

Chromatography separates analytes on the basis of different physiochemical properties. Reverse-phase liquid chromatography (RPLC) remains predominant for fractioning complex peptide mixtures (*16*). This technique separates by differences in polarity. A nonpolar stationary phase and a polar mobile phase create a phase boundary. Analytes that are more hydrophobic take longer to travel through the column in the mobile phase because they will enter the stationary phase more often and remain there longer. A mobile phase gradient is one that gradually changes composition over time, washing different analytes off the column as the percentage of polar solvent decreases. Why is this separation so useful to perform before MS analysis? A low abundance protein gains more representation through digestion, increasing its chances for detection, but those peptides need to be more intense than others (and background noise) during an MS scan to be selected for isolation and fragmentation in data-dependent MS/MS (*14*). In other words, entering the MS in a fraction of fewer peptides increases lower abundance peptides' likelihoods of detection and identification. It also increases the total number of proteins matched to the peptide spectra. Additionally, as an in-line method, high performance liquid chromatography (HPLC) fractions are fed directly into the MS by electrospray ionization (ESI), minimizing sample loss. This ionization source suffers from ionization suppression if there are too many contaminants or peptides spraying at once (*14*). HPLC aides in effective ionization by creating distinct peaks with fewer analytes entering the source simultaneously (*14*). Improving the resolution and peak capacity (a columns'

maximum number of peaks separated per gradient time) results in higher sensitivity and more well-resolved spectra (*3, 14, 16*). This can be achieved multiple ways.

HPLC operates under high pressures in order to increase the efficiency of separation and speed of analysis; it can be automated in this way. Smaller internal diameters improve sensitivity, as well as longer columns (*3, 14, 16*). As inner diameters decrease, so must the flow rate; therefore, for proteomics, nano HPLC (or UPLC) systems paired to nanospray ionization (NSI) are incredibly desirable (*3, 14, 16*). In HPLC, flow-splitting is required to achieve nL/min flow rates, but ultra high pressure liquid chromatography (UPLC) avoids flow-splitting, utilizing pressures over 1000 bar (*3*). UPLC allows for the use of smaller particles in packed columns, reducing peak widths, and improving sensitivity further (*3, 14*).

Multidimensional liquid chromatography (MDLC) can, alternatively, increase peak capacity by placing multiple orthogonal separations in a row (*3, 14, 16*). MDLC consumes more sample, but an online approach, one in which each fraction is transferred directly from one column to the next, helps preserve sample and maintains high-throughput (*16*). Online mode requires chromatographic compatibility in terms of mobile phase or separation technique, limiting flexibility (*16*). Additionally, the buffers should be compatible with ionization technique (*16*). Most commonly, ion exchange chromatography is performed prior to RPLC (*3, 16*). Both cation and anion exchange have been performed in-line with RPLC prior to MS introduction (*16, 20, 21*). For example, SCX-RP has been accomplished with biphasic columns (*20*) and by multistep salt-injections and column-switching (*21, 22*). Of significant importance in these techniques, and most proteomic analyses, are trapping columns that aid in desalting (*3*). Trap columns, or precolumns, help "trap" your sample, while washing away unwanted buffers or contaminants to limit ionization suppression in electrospray sources and reduce sample complexity. For example, a TiO_2 trap

column can be used for the enrichment of phosphopeptides (*3*). In work performed by Kuroda et al., peptides were washed from a titanium trap column onto a silica trap column, the buffers were washed away, and then the peptides were transferred to an RP separation column in-line with MS (*23*).

Once the sample reaches the MS, it must be ionized and transferred to the gas-phase. There are two main ionization methods used for proteomic analysis: matrix assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) (*8, 14*). Minimal fragmentation occurs during these processes making them soft ionization sources. For MALDI, the sample is mixed into a matrix which can absorb laser light. The sample dries with the matrix as a crystalline solid on a plate (*8, 24*). A laser is shone onto the matrix, the surface heats rapidly, and the sample is vaporized along with the matrix (*24, 25*). During this laser ablation, the matrix transfers protons to the sample molecules. MALDI creates mostly singly-charged analyte ions necessitating a mass analyzer with high mass range. MALDI is generally used for less complex peptide or protein mixtures (*8*), and is frequently performed after gel separations (*14*). MALDI typically requires a large amount of sample preparation, but it has been accomplished directly from 2D gels (*26*). It can follow HPLC, but requires fraction collection and transfer to MALDI plates (*14*). It has a higher resistance to surfactants during ionization when compared to ESI, but ESI is most often coupled to LC because of its ability to ionize sample directly from liquid solution, in-line with the chromatographic separation (*8, 14*).

In ESI, the sample in solution is passed through a metal capillary held at a high electric potential, and charge accumulates at the air/liquid junction (*27*). A positive potential will accumulate positive charge, and a negative potential will accumulate negative charge (*27*). For the formation of positively-charged ions, solvents contain a small percentage of acid, like formic acid

for example, that transfers protons to the analytes. The electric field, coulombic repulsion, and surface tension result in the formation of a Taylor cone, a conical spray exiting the capillary (*27*). Droplets are ejected from the cone as charge repulsion overcomes surface tension (*27*). As solvent evaporates from the droplets, coulombic repulsions destabilize the droplet again, resulting in fission (*27*). There is a cycle of fissions that results in smaller and smaller drops until they are small enough for the analyte to be transferred to the gas-phase, charged from solution (*27*).

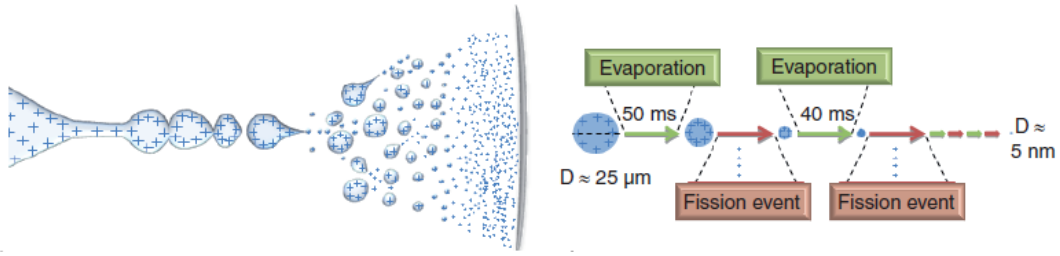

Figure 1.2 Electrospray Ionization Process (*27*)

Sample molecules can suffer from ionization suppression in ESI because other molecules in solution can compete to hold the excess charge (*27*). For proteomics, nanospray ionization (NSI) is frequently chosen over ESI because it can accommodate the low flows of efficient nano-HPLC and is more resistant to ionization suppression. NSI operates by the same principles of ESI, but the emitters have smaller inner diameters, resulting in its ability to spray from nL/min flow rates and tolerate salt contamination (*28, 29*). The smaller initial droplets yield a higher charge-to-volume ratio, prompting ion release after fewer generations of fissions and a lower salt concentration increase per fission (*28, 29*). By this, the ionization efficiency is higher in NSI than ESI (*28, 29*).

After ionization, the peptides enter the mass spectrometer. Several mass analyzers are useful for bottom-up proteomic analysis, but the linear ion trap (LIT) is most commonly employed (*14*). While all mass analyzers operate to separate ions by their mass-to-charge ratios (m/z), they execute this task by different means (*14*). A LIT confines ions by a two-dimensional radio

frequency field and by stopping potentials applied to end electrodes, allowing for both isolation and fragmentation (*14, 30*). LITs are useful for tandem MS in time, serving mainly to identify proteins or quantify without labeling (*14*). For data-dependent mass spectrometry, an initial scan takes place to identify abundant peptide ions, then the LIT traps an identified precursor ion of interest by ejecting all ions but those within the specified m/z window (*14*). Next, the isolated ions will undergo fragmentation, and the product ions are scanned to create a product ion spectrum (*14*). The process is repeated many times during a sample analysis.

Other useful mass analyzers include the quadrupole (Q), Time-of-Flight (ToF), Orbitrap, and FT-ICR (*14*). Interfacing multiple mass analyzers can increase experimental capabilities. For example, a drawback to LIT instruments is their lower mass accuracy (*8*). Combining a LIT with an Orbitrap will improve resolution and mass accuracy, increasing confidence in precursor ion measurements, protein identification, and quantification (*14*). Orbitrap and FT-ICR (Fourier transform ion cyclotron resonance) instruments both use Fourier transform to ultimately convert frequency of ion motion to m/z data resulting in high resolving power. While these FT instruments are incredibly powerful, their cost limits their widespread use. The FT-ICR has the highest mass accuracy of mass analyzers, but has complex operation and low fragmentation efficiency (*8*, *14*). The Orbitrap is faster and more sensitive, and less costly to maintain (*14*, *16*). The Orbitrap commonly serves to identify proteins and post-translational modifications and to perform isotope-labeled quantification (*14*). Top-down proteomics necessitates the use of these high-resolution instruments for the sample complexity and large size of intact proteins (*16*). The rToF has relatively high mass accuracy as well and theoretically unlimited mass range (*14*, *16*). A triple quadrupole (QQQ) instrument performs tandem MS in space, has notable dynamic range and sensitivity, and is used in targeted protein quantification (*14*). A QToF combination mimics the

QQQ setup, but provides higher mass accuracy because the ToF detects precursor or fragment ions (*14*). The ions entering a ToF are accelerated to have the same kinetic energy ($1/2mv^2$), then they pass through a drift region to the detector. Ions of different mass reach the detector at different times because they have differing velocities. The ToF does not scan like other mass spectrometers but identifies ions' m/z ratios by the time it takes them to hit the detector (*31*). Ultimately, a choice of mass analyzer reflects the goals of the proteomic analysis, and is informed by differing needs for speed, accuracy, and sensitivity (*14*).

Table 1.2 Comparison of Mass Spectrometers (*14, 16*)

Instrument	Resolving Power	Mass Accuracy	Dynamic Range	Common Proteomic Applications
LIT (LTQ)	2,000[a]	+[b]	+	Bottom-up
QQQ	3,000	++	+++	Bottom-up quantitation, target quantitation (SRM, MRM, PTM monitoring)
LTQ-FTICR	600,000[a]	++++	+	Bottom-up, top-down, PTM characterization, quantitation
LTQ-Orbitrap	100,000 - 240,000	+++	+	Bottom-up, top-down, PTM characterization, quantitation
Q-ToF	22,500 - 100,000	+++	++	Bottom-up, top-down, PTM characterization, quantitation
Q-Orbitrap	140,000	+++	+	Bottom-up, top-down, PTM characterization, quantitation (PRM)

a – mass resolution achieved at normal scanning rate, but higher resolution can be accomplished at slower speeds
b – "+" indicates performance. The more "+," the better the performance in that aspect

Within the mass analyzer, a peptide can be made to undergo fragmentation. The main fragmentation methods are collision-induced dissociation (CID), electron capture and electron transfer dissociation (ECD, ETD), and photon-based dissociation like UV photodissociation (UVPD) and infrared multiphoton dissociation (IRMPD) (*14, 16*). CID is most commonly employed for bottom-up approaches and arises from collisions of the analyte with background gas like helium or argon. It produces mainly b- and y-ions which result from the cleavage of the amide

bond and the retention of positive charge on either the N-terminal or C-terminal fragment, respectively (*14*). Other labile bonds, like phosphate bonds, will also be cleaved, eliminating some PTM information (*16*). Gas-phase reactions with electrons can generate random fragmentation of the peptide backbone; this is the principle of electron-based dissociation methods like ECD and ETD. ECD uses thermal electrons, and ETD uses an anionic electron donor, but both result in c- and z-ions (*14*). Phosphate groups are kept intact during ECD or ETD, offering an advantage over CID for PTM studies (*14*). A disadvantage is this technique's dependence on the charge state of the peptide or protein; a precursor ion must at least be doubly-charged to be fragmented by ECD/ETD (*14*). In IRMPD, IR laser light is shone on ions, and they absorb multiple photons which increase their vibrational energy until fragmentation results, producing similar ions to CID (*16*). IRMPD is easily accomplished with CO_2 lasers and modest modifications to the mass spectrometer. UVPD is similar, but results in extensive fragmentation, leading to highly complex product ion spectra. ECD/ETD and UVPD are chosen often for top-down proteomics because they can more efficiently fragment large proteins (*16*).

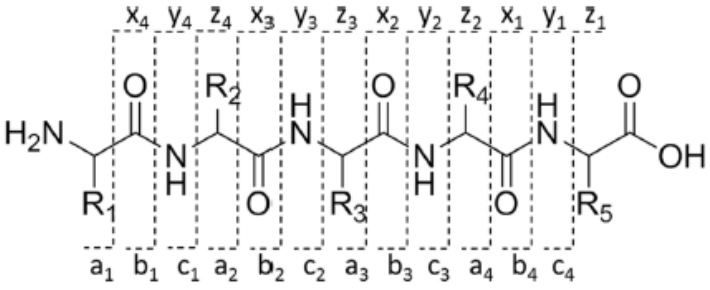

Figure 1.3 Fragment Ions (*14*)

Upon completion of MS analysis, the spectra are deciphered, and the peptides are identified. Not all proteomic experiments apply tandem MS/MS. Some MS setups, like MALDI-TOF, use high resolution to measure the m/z of intact peptides. Peptide mass fingerprinting is used

to identify these peptides and their protein of origin (*8*). Using the experimental peptides' masses, a protein database is searched (*8*). The masses will be matched to the calculated peptide masses for the protein in the database (*8*). To perform this kind of correlation with confidence, the sample needs to be a purified target protein (*8*).

For MS/MS studies, the fragmentation spectra and precursor mass are used to match an amino acid sequence that can then be matched to a protein-of-origin. The immense amount of spectral data for a shotgun analysis requires that software be utilized to accomplish this task. De novo sequencing, estimating a sequence directly from the spectra, is challenging, so searching algorithms utilize sequenced protein databases to match fragmentation spectra to a peptide sequence and a protein (*8*). There are multiple approaches to database searching. A peptide sequence tag approach extracts a short amino acid sequence from the fragmentation peak pattern and acts as a probe to find the peptide's origin (*8*). The cross-correlation method creates theoretical spectra for peptide sequences in the database, then scores the experimental spectra against them to match a peptide sequence (*8*). In the probability sequence matching approach, fragment ions are predicted from sequences in the database, and these are matched to the experimental fragment peaks, yielding a comparison score (*8*). While none of the searching programs are perfect, and they of course require interpretation from a researcher, they enable the analysis of large amounts of proteomic data.

In comparison to top-down proteomics, bottom-up is performed on a much wider variety of instruments; it doesn't require such high mass range or spectral resolution. In addition, peptides are more easily separated, ionized, and fragmented than intact proteins (*14*). The array of separation techniques applicable to peptides enables bottom-up approaches to tackle larger numbers of sample components than top-down, so whole protein approaches are more frequently

applied to investigate single proteins (*14, 16*). Bottom-up investigations have drawbacks as well, some of which can be addressed by top-down examinations. Peptide redundancy between proteins can be a problem when identifying a peptide's protein-of-origin, but, by analyzing whole proteins, that issue is completely avoided (*14*). Sequence information and alternative splicing variations can also be revealed over the course of a top-down experiment (*14*). Sequence coverage can be incomplete in bottom-up proteomics (*16*). Proteolytic cleavages of peptides can result in fragments that are too big to be detected by common MS instrumentation (*16*). Additionally, combinatorial PTMs are nearly impossible to confirm when studying proteins as smaller fragments (*16*), and some PTMs may even be lost during bottom-up sample preparation. Despite their shortcomings, both approaches offer unique information regarding cellular composition and operation. Combining these complementary techniques provides a fuller view of the proteins under investigation.

1.2 *Helicobacter pylori*

H. pylori is a gram-negative, spiral-shaped bacterium that settles in the harsh environment of the human stomach (*32*). Today, about 50% of the world's population hosts them (*33*), which is great cause for worry because *H. pylori* is no harmless symbiont; infection is the greatest known risk factor for gastric cancer (*34*). Its presence causes chronic gastritis (noticed or unnoticed) and increases risk of duodenal and gastric ulcers, in addition to gastric cancer (*34*). Gastric cancer is one of the leading causes of cancer-related deaths around the globe (*34*). To add to this alarming carcinogenic ability, these bacteria are also very persistent; they can survive for the lifetime of the host, evading the immune response and the acidity of the stomach (*33, 34*). With increasing antibiotic resistance, it of utmost importance to study their survival mechanisms to find new targets for drugs to treat and eradicate infection (*35*).

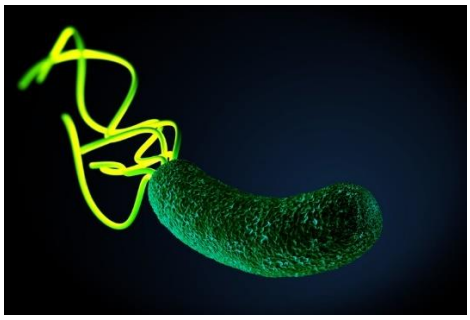

Figure 1.4 *Helicobacter pylori*

These bacteria live within the gastric mucosal layer and adhere to gastric epithelial cells via adhesin proteins (*34, 36*). The expression and repression of these adhesins could play a key role in persistence of infection. SabA, sialic acid binding protein, represents one of these adhesins. The gene that codes for this protein, sabA, is repressed by the phosphorylated ArsR response regulator protein in acidic environments via the ArsRS signaling system (*36*). Interestingly, the lab of Dr. Mark Forsyth has found that sabA repression at acidic pH is ArsS dependent, but phosphorylated ArsR independent. ArsS is a histidine kinase that phosphorylates ArsR in this two-component signaling system. In the wild-type strain, sabA transcription is repressed at pH 5 compared to pH 7 conditions. When the arsS gene was deleted, this repression did not occur, but when the canonical phospho-accepting site, aspartic acid 52 (D52) was mutated to glutamic acid (E) or asparagine (N) and arsS remained un-mutated, repression of sabA transcription still occurred at pH 5 and was slightly repressed at pH 7. These mutants were not able to be phosphorylated at the 52nd position, yet sabA was still repressed at levels similar to wild-type at acidic pH. This leaves some questions. Providing explanation for the minor repression at pH 7, are these mutants taking a phosphomimetic shape? Are they being phosphorylated at other positions, perhaps D47 or D59? Is there cross-talk between ArsS and non-cognate response regulators? Is ArsR

undergoing other post-translational modification (PTM) mediated by ArsS, like acetylation, that allows sabA repression?

To address questions regarding post-translational modification of wild-type and mutant ArsR at neutral and acidic pH, bottom-up proteomics was employed. Preliminary investigations in which purified ArsR protein was prepared for LC-MS analysis by digestion with trypsin and Lys-C and trypsin and Asp-N proteases are discussed here. The aspartic acid residues of interest, D47, D52, and D59 were detected and identified in SEQUEST. Shotgun proteomic analysis of *H. pylori* (26695) cultures was also performed. With electrospray serving as the ionization source, it was determined that ArsR is not abundant enough in a shotgun sample to perform PTM analysis. Methods for nanospray and electrospray ionization were progressed.

Chapter 2: Experimental Methods

2.1 Sample Preparation

2.1.1 Purified ArsR

Dr. Mark Forsyth provided the purified ArsR (response regulator ompR, HP_0166 gene, strain 26695) samples of concentration 7ug/ul in 10Mm Tris (pH 7.5), 50mM KCl, and 1mM DTT, eluted originally from an affinity column in 50mM NaH_2PO_4, 300mM NaCl, and 250 mM imidazole with buffer exchanged using a microconcentrator. These ArsR samples contained a histidine 13 epitope tag on the c-terminus. These were produced in Dr. Forsyth's lab by creating mutagenic oligonucleotides that add 4 alanine codons in the location of the native stop codon, followed by 13 tandem histidines codons, and a stop codon.

The Pierce Mass Spec Sample Prep Kit for Cultured Cells, Pierce C18 spin columns, additional cell lysis buffer, and Asp-N and additional trypsin MS grade proteases were purchased from Thermo Fisher Scientific. The Pierce Mass Spec Sample Prep Kit for Cultured Cells contains cell lysis buffer, digestion buffer, no-weigh dithiothreitol (DTT) tubes, iodoacetamide (IAA) single-use tubes, trypsin storage solution, Pierce digestion indicator, and MS-grade Lys-C and trypsin proteases. Ultrapure water was obtained via a Milli-Q ultrapure water system. Acetic acid, acetone, acetonitrile, trifluoroacetic acid, methanol, and zinc acetate were manufactured by Fisher Scientific. Tris HCl and formic acid were acquired from Sigma-Aldrich.

Initially, concentrated ArsR samples were digested and analyzed to assess protease cleavages and protein coverage. Samples were diluted with cell lysis buffer from 7 µg/µL to 1 µg/µL since the procedure was optimized for 100 µg of protein at 1 mg/mL. The protocol can accommodate samples between 10 and 200 µg at concentrations between 0.2 mg/mL and 1 mg/mL by adjusting the amounts of reagents used. The first sample contained 70 µg of protein, the following samples each contained more than 100µg, but only 100µg were used (about 14.3 µL). The cell lysis buffer was prepared by dilution of the 20x concentrated sample in ultrapure water (5 mL of concentrated buffer in 100 mL of water).

After dilution with cell lysis buffer, the samples were reduced and alkylated to prevent disulfide bridges from forming. Reduction by DTT was performed first. 500 mM DTT solutions were made immediately before use by puncturing the foil of a no weigh DTT tube with an empty pipette tip, adding 100 µL of ultrapure water, and pipetting up and down to dissolve the contents of the tube. DTT solution was added to the sample to produce a final DTT concentration of about 10 mM. For 100 µL of sample solution, 2.1 ul of DTT solution was added. After mixing and incubating at 50˚C for 45 minutes, the samples were cooled for 10 minutes at room temperature

and then subjected to alkylation. The 500 mM IAA solutions were also made immediately before each use. A single-use tube was punctured by a clean pipette tip, and 100 µL of cell lysis buffer was added to the tube, pipetting up and down to dissolve and mix. The solution was protected from light. Enough solution was added to the sample to produce a final concentration of about 50 mM. For a sample of about 100 µL (DTT has already been added), 11.2 µL of IAA solution would be used. After mixing and incubating at room temperature for 20 minutes protected from light, 460 µL of chilled (-20˚C) 100% acetone was added in four volumes (115 µL each). The Eppendorf tube containing the sample was then vortexed and allowed to incubate overnight at -20˚C to precipitate protein.

Next, the sample was centrifuged at about 16,000 x g and 4˚C for 10 minutes using either Dr. Lizabeth Alison's 5418 R Eppendorf centrifuge or Dr. Forsyth's Hettich Universal 320R, the supernatant was carefully removed without dislodging the pellet. Next, 50 µL of chilled (-20˚C) 90% acetone was added. The pellet was resuspended by vortex mixing, and the sample underwent centrifugation by the same conditions as previously mentioned, this time for 5 minutes. The acetone was removed, and the pellet was allowed to air dry for only 2-3 minutes before proceeding to digestion.

100 µL of digestion buffer was pipetted onto the pellet. Resuspension was achieved by pipetting up and down. During the first use of the Pierce Mass Spec Sample Prep Kit, 40 µL of ultrapure water was added to the bottom of the Lys-C vial. This was incubated at room temperature for 5 minutes before use. Lys-C was dissolved by pipetting up and down. Enough Lys-C was added to reach a 1:100 enzyme-to-substrate ratio. For example, 2 µL (1 µg) was added to a 100 µg protein sample. A sample was then mixed and incubated for 2 hours at 37˚C. The remaining Lys-C was stored at -80˚C between uses. Upon first use, trypsin was also reconstituted. 40 µL of the trypsin

storage solution was added to the bottom of the vial, which was incubated for 5 minutes. Pipetting followed to dissolve. Additional MS-grade trypsin was reconstituted with 20 µL of 50mM acetic acid. Trypsin was added in a 1:50 enzyme-to-substrate ratio. For example, 4 µL (2 µg) was added to a 100 µg sample. Each sample was incubated at 37˚C overnight. Samples were frozen at -80˚C the next day to stop digestion. Finally, the sample underwent gentle evaporation by speed vacuum (Thermo Scientific Savant DNA 120 SpeedVac Concentrator) and was resuspended in 400 µL of HPLC solvent A (98% ultrapure water, 2% ACN, 0.2% formic acid). With Lys-C and trypsin digestions, C18 spin column de-salting was not required, but a PVDF syringe filter was used to clean-up these samples.

Digestion was altered for the last 100 µg purified sample to include Asp-N and exclude Lys-C. The digestion steps began by dissolving the dried protein pellet in digestion buffer, as stated above. Immediately, a minimal trypsin digestion was performed. The enzyme-to-substrate ratio was lower; 1-2 µL of trypsin solution was added, and incubation overnight occurred. The digestion buffer was evaporated, and the sample was resuspended in 20 µL of the Asp-N 5X digestion buffer (250 mM Tris·HCl with 2.5 mM zinc acetate) and 50 µL ultrapure water. The buffer was prepared as a 100 mL stock solution containing 3.922 g of Tris·HCl and 0.0466 g of zinc acetate. The protocol was written for a 50 µg sample, so the 100 µg sample was resuspended in double the water and digestion buffer volumes written. Following resuspension, the sample was halved; 35 µL was used for the first digestion, and 35 µL was stored at -80˚C. The volume of the sample was then adjusted to 50 µL with ultrapure water. The vial of 2 µg of lyophilized Asp-N was reconstituted the first use. The protocol suggested a reconstitution concentration of 1mg/mL and a 1:20 enzyme-to-protein ratio. Considering the high cost of the protease and the difficulty of working with such low volumes, Asp-N was reconstituted with 4 µL of ultrapure water creating a

0.5 mg/mL concentration. The digestion was performed at a 1:200 enzyme-to-protein ratio with 0.5 µL (0.25 µg) of Asp-N solution to start. The sample was incubated overnight at 37˚C. Next, the sample was de-salted via a C18 spin Column, the solvent was evaporated by speed vac, and the sample was resuspended in 400 µL of solvent A. The second 35 µL of peptide was digested with Asp-N in the same manner, but was not de-salted immediately following. The solvent was evaporated by speed vac, and an additional trypsin digest was performed to improve protein sequence coverage. The sample was resuspended in 50 µL of trypsin/Lys-C Digestion Buffer, and 2 µL of trypsin solution was added. It was incubated overnight and then de-salted by C18 spin column, speed vacuumed, and resuspended in 200 µL of solvent A. The spin column procedure can be summarized by the image below.

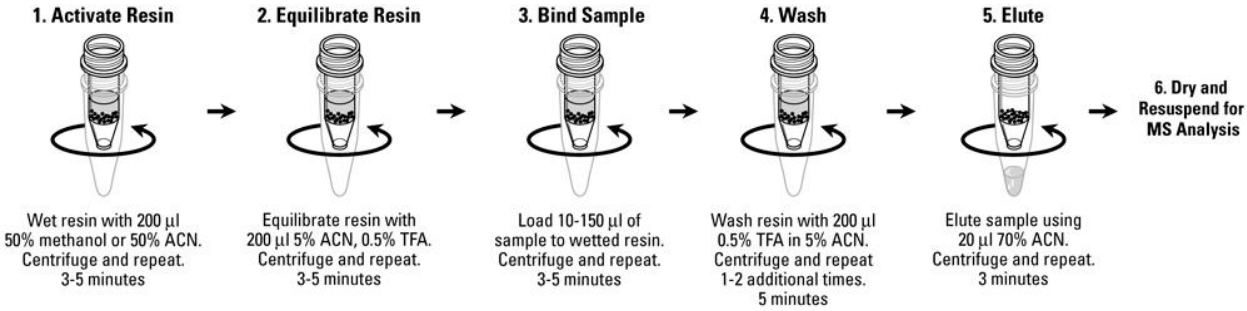

Figure 2.1 Spin Column Protocol Summary (*37*)

The activation solution, equilibration/wash solution, sample buffer, and elution buffer were each prepared in 10mL volumes. The activation solution was made to contain 50% methanol, the equilibration/wash solution was made to contain 0.5% TFA in 5% ACN, the sample buffer 2% TFA in 20% ACN, and the elution buffer 70% ACN.

This procedure was the same for both for each of the 50 µg digested purified protein samples. First, sample buffer was added in a 3:1 sample to buffer ratio (about 17 µL). The column was tapped to settle the resin, the ends were uncapped, and it was placed in a receiver tube (1.5

mL Eppendorf tube). 200 µl of activation solution was added to rinse the walls of the column and wet the C18 resin. The column was centrifuged resting in the receiver tube in an Eppendorf minispin centrifuge at about 1500 x g (around 3500 rpm) for 1 minute. The flow-through was discarded. The activation solution addition and centrifugation was repeated, discarding flow-through. Next, 200 µL of equilibration/wash solution was added. The column was centrifuged again at 3500 rpm for 1 minute, and this flow-through was discarded. The equilibration/wash was repeated, discarding flow-through. Next, the sample was loaded on top of the resin bed with the column placed in a new receiver tube. The sample was centrifuged under the same conditions as previously. The flow-through was recovered and placed back on top of the resin bed, repeating these steps to ensure complete sample binding. The Eppendorf tube containing this flow-through was saved for both samples in case of ineffective sample binding. Next, the column was washed with 200 µL of equilibrations/wash solution in a new receiver tube and centrifuged under the same conditions as the prior steps. The flow-through was discarded and the wash repeated, still discarding flow-through. Finally, the sample was eluted into a new receiver tube by the addition of 20 µL of Elution Buffer on top of the resin bed and centrifugation at 3500 rpm for 1 minute. This step was repeated with the same receiver tube. The eluents containing the samples were dried by speed vacuum. The retained flow-throughs from sample binding was also speed-vacuumed. The first sample was resuspended in 400 µL of solvent A. The second sample was resuspended in 200 µl of solvent A. The first flow-through was reconstituted in 200 µL of solvent A and filtered by a PVDF syringe filter. The second flow-through was reconstituted in 200 µL of solvent A, but not filtered with a syringe filter. After this de-salting and filtration, these samples were ready for HPLC-MS/MS analysis.

2.1.2 Shotgun Proteomics

Dr. Mark Forsyth cultured the *H. pylori* (strain 26695) cells on plates, and provided enough cells for use in the shotgun procedure. Lyses were performed in his lab under his supervision for biosafety. The materials used for shotgun sample preparation are the same as for the purified protein sample preparation with the addition of The Pierce BCA Protein Assay Kit and phosphate-buffered saline (PBS). The Pierce BCA Assay Kit was purchased from Thermo Fisher Scientific and includes BCA reagent A (sodium carbonate, sodium bicarbonate, bicinchoninic acid and sodium tartrate in 0.1 M sodium hydroxide), BCA reagent B (4% cupric sulfate), and albumin standard ampules (bovine serum albumin 2mg/mL in 0.9% saline and 0.05% sodium azide). The PBS was purchased from Sigma Aldrich. The PBS was manufactured by Boston BioProducts.

To prepare a shotgun sample, enough cells are lysed to try to harvest at least 100 µg of protein. As stated previously, the procedure is optimized for 10-200 µg at a 1 mg/mL concentration. Cells are washed and lysed with cell lysis buffer. DNA is broken up using a micro-sonicator, and the sample is centrifuged to obtain the cellular proteins. Below is an image summarizing the shotgun proteomics procedure, including the steps after cell lysis. An example lysis procedure follows.

Figure 2.2 Shotgun Proteomics Procedure Summary (*13*)

Dr. Forsyth cultured *H. pylori* cells on plates and scraped enough off to complete the procedure optimally, placing the cells into a 1.5 mL Eppendorf tube. 100 µL of PBS was added to the tube, and the sample was centrifuged at less than 1000 x g to pellet the cells. The supernatant

was removed, and 100 µL of cell lysis buffer was added, pipetting up and down to break up the cells. Incubation occurred in a hot water bath at 95˚C for 5 minutes. The Eppendorf tube was then placed on ice for 5 minutes to cool the sample. Following, the sample was micro-sonicated on ice to shear DNA. The micro-sonicator was placed in the sample for 30 seconds and then allowed to cool in the ice; this was repeated twice more. Finally, the lysate was centrifuged at 16000 x g for 5 minutes at 4˚C in a Hettich Universal 320R, and the supernatant containing the proteins was transferred to a new tube and stored at -20˚C until the BCA Assay was performed to determine protein concentration.

The BCA Assay can be prepared using test tubes or a microplate. With the availability of Dr. Douglas Young's microplate UV-Vis, a BioTek Synergy HT microplate reader, the amount of sample spent to determine the concentration was reduced in comparison to a traditional UV-Vis instrument; therefore, the shotgun samples were assessed using the 96-well microplate procedure. To create a calibration curve from which unknown protein concentration could be interpolated, standards of known concentration were made using albumin protein. Standard solutions were made according to the table below.

Table 2.1 BCA Assay Standards

Vial	Volume of Ultrapure Water (µL)	Volume (µL) of and Source of Albumin	Albumin Concentration (µg/mL)
A	0	300 of Stock	2000
B	125	375 of Stock	1500
C	325	325 of Stock	1000
D	175	175 of vial B dilution	750
E	325	325 of vial C dilution	500
F	325	325 of vial E dilution	250
G	325	325 of vial F dilution	125
H	400	100 of vial G dilution	25
I	400	0	0, Blank

Next, the working reagent was prepared by combining 2.5 mL of reagent A with 50 µL of reagent B (50:1 ratio). 25 µL of each standard were placed into wells of a transparent 96-well microplate. 25 µL of the protein cell lysate was also placed into a well. Next, 200 µL of working reagent was pipetted into each well containing a standard or sample.

A lid was placed on the microplate and it was placed into the microplate UV-vis. A protocol was created in Gen5 by BioTek, the microplate reader software, and saved as "BCA Assay." The plate was first shaken for 30 seconds. Next, the instrument's temperature was brought up to 37˚C, and the plate was incubated for 30 minutes. It was allowed to cool to room temperature, and absorbance readings were then taken for each well at 562 nm. This data was recorded and exported to excel for interpolation of the standard curve and determination of the protein concentration.

Once the protein concentration of a cell lysate is determined, one can proceed to the reduction, alkylation, and precipitation steps and then through digestion. Proteolysis was performed in the same way as the purified ArsR sample preparation states above, section 2.1.1. Initial shotgun proteomic analysis was performed after digestion with Lys-C and trypsin. Asp-N was not used while evaluating the procedure.

2.2 High Pressure Liquid Chromatography (HPLC)

Reverse Phase Liquid Chromatography (RPLC) was performed on a Shimadzu UPLC instrument. 100 µL of samples were pipetted into HPLC vials and placed into the sample tray. The autosampler took 10 µL of sample, injecting it into the sample loop. The sample was run through the HPLC column, an ACE Excel 3 Super C18 heated with column oven (40˚C), at 400 µL/min with a time-dependent binary gradient flow shown in the table and curve below. A 100-minute gradient was used.

Table 2.2 100-Minute Direct Gradient, ESI

	Time (min)	Event	Parameter
1	0.01	Start	
2	0.02	Total Flow	400 µL/min
3	0.1	Pump B Concentration	2%
4	5.02	Pump B Concentration	2%
5	60.00	Pump B Concentration	55%
6	60.10	Pump B Concentration	90%
7	70.00	Pump B Concentration	90%
8	70.10	Pump B Concentration	2%
9	99.10	Total Flow	400 µL/min
10	100.00	Stop	

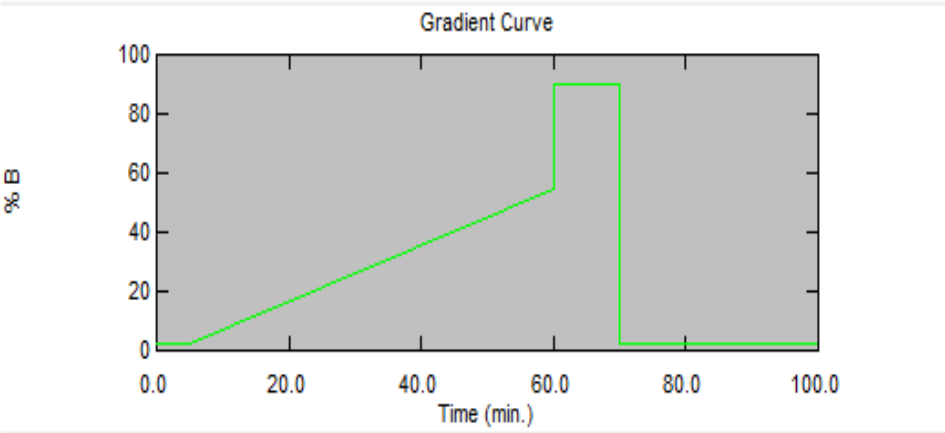

Figure 2.3 HPLC Gradient Curve 100-Minute Direct Method, ESI

Solvent A was comprised of 98% ultrapure H_2O, 2% ACN, and 0.2% formic acid. Solvent B was comprised of 98% ACN, 2% ultrapure H_2O, and 0.2% formic acid. These were prepared as needed in 1000mL batches. Differential pumping produced the varying solvent percentages over time. The sample flowed directly into the electrospray ionization (ESI) source to enter the MS for analysis. A blank of solvent A was run between each sample. Injection volume for each blank was also 10 µL and the same 100-minute method was used.

2.3 ESI and Data Dependent Mass Spectrometry

Mass spectrometry was performed on a Thermo LTQ-XL linear ion trap mass spectrometer. ESI spray voltage was set to 5kV for all runs, with the ESI probe set to position B.

Auxiliary gas flow rate was set to 60 arbitrary units. Sheath gas flow rate was always set to zero. The heated capillary temperature was 275˚C.

The XCalibur program saved and executed MS methods. The LTQ first recorded a mass spectrum by scanning from 350 m/z to 1700 m/z. The program then selected the four most intense peaks for isolation and fragmentation in separate scan events, making a total of 5 events. The product fragmentation spectra were each saved by XCalibur. This full scan, isolation, and fragmentation process was repeated for the entirety of the HPLC gradient. CID was the mode of fragmentation; collisions with background helium gas in the trap caused dissociation of selected ions. The default charge state was 2, the normalized collision energy was 35%, the activation Q was 0.250, the activation time was 30 ms, and the isolation width was 4 m/z. Data was acquired in centroid mode. If the same mass reappeared within 30 seconds, it was then added to a dynamic exclusion list and excluded for 60 seconds. Additionally, some contaminant peaks were always excluded. Many of these were plasticizers and polysiloxanes. The excluded m/z were: 195.00, 278.10, 370.90, 371.00, 372. 373, 388.13, 429.09, 445.12, 462.15, 503.11, 519.14, 524.00, 550.63, 593.15, 606.10, 625.39, 667.18, 679.41, 695.43, 701.39, 741.20, 753.48, 1022.20, 1122.20, 1222.20, 1322.20, 1422.20, 1522.20, 1622.20, 1722.20, and 1822.20 with a 1.5 m/z exclusion width.

2.4 SEQUEST

SEQUEST is a bioinformatic database searching program used to interpret the proteomics data acquired from fragmentation spectra; it aids in peptide and protein identification from tandem MS/MS spectral data. It functions by matching the measured mass of peptide ions, considering a mass tolerance and PTMs specified by the user, to amino acid sequences in a protein database, then predicting the fragment ions for each matched sequence (*38, 39*). The protein databases are

input in the form of FASTA files containing primary protein sequence information represented by one-letter code, and the user's input activation type helps dictate which kind of fragment ions will be predicted (*38, 39*). Then, a preliminary score, the SpScore, is generated by matching predicted fragment ions to those in the experimental spectra (*38-40*). If these matches pass the SpScore filter, the program generates theoretical spectra for matched amino acid sequences using predicted fragment ions' mass-to-charge (m/z) ratios and predicted magnitude components (*38*). A cross-correlational scoring value is calculated which compares the theoretical spectra to the experimental spectra (*38-40*). After matching the theoretical spectra to the experimental spectra, and producing cross-correlation scores for the peptides, the program also indicates the proteins from which the matched sequences were derived (*38-40*). A protein score is presented based on the confidence of the peptide spectral matches and how many peptides were matched from that protein (*40*). SEQUEST helps evaluate mass amounts of spectral data by identifying possible peptides and proteins from the sample and aiding in the confirmation of PTMs.

To begin, the files that have been saved by XCalibur from an LC/MS run are saved to the computer hosting the SEQUEST program. "Thermo Proteome Discoverer 1.4.1.14" is opened. This is a software that supports the SEQUEST database searching algorithm. The workflow editor is selected, and a template workflow is opened, or a new workflow is created. The workflow should include a spectrum file selection, a spectrum selector, a sequence database search, and a decoy database search, but can include enhanced filtering and spectral processing, spectral library searching, peptide spectral match validation, and post-translational modification analysis as well. In the workflow editor window, parameters are input for each step of the workflow. The Xcalibur files are input under spectrum file selection. In the spectrum selector step, the program sorts through incoming spectra. Spectrum selector parameters identify the kind of spectra that will be

analyzed. For example, this is where one would indicate the MS order and the activation type. This gives the program information about incoming data and specifies the precursor for higher order MS. During the sequence database searching step, precursor ions' masses and PTMs are matched to amino acid sequences in the protein database, fragment ions are predicted for the sequences and matched to experimental fragment ions, and theoretical spectra are generated and matched to experimental spectra. Parameters for this step include inputting a FASTA file, specifying a type of digestion, and indicating peptide modifications and mass tolerances. The decoy database search helps estimate the number of false positive matches resulting from the true database search. Parameters involve a strict and relaxed false discovery rate.

The workflow optimized for data analysis was saved as a template. This template included spectrum file selection, a spectrum selector, a sequence database search, and a decoy database search. Spectrum selector parameters included: precursor mass range of 300 Da to 5000 Da, MS order of MS2, and CID activation type. SEQUEST was used as the searching program. An *H. pylori* (26695) proteome FASTA file was the default protein database saved in the template. FASTA files were saved from the UniProtKB/Swiss-Prot. The default enzyme name was Trypsin (Full) but Asp-N and "no enzyme" were also selected for some workups. Only certain protease pairings could be selected for the enzyme parameter. Trypsin and Asp-N had to be selected separately. The precursor and fragment mass tolerances were chosen carefully after systematic work-ups of the same data. High peptide Xcorr values were considered in combination with high percent coverage to select these tolerances. The precursor mass tolerance was selected to be 1.5 Da and the fragment mass tolerance was 0.6 Da. The table below illustrates the work-ups leading to these choices.

The following sequence database search parameters were also specified: a maximum of 2 missed cleavage sites, a 5- or 6-144 peptide length range, b- and y-type ions calculated, a static modification of cysteine carbamidomethylation (+57.021 Da), and dynamic modifications of N-terminal acetylation (+42.011 Da), C-terminal oxidation (+15.995 Da), and methionine oxidation (+15.995 Da). Both b and y ions are commonly created during low-energy CID. The static modification of cysteine is produced by protein reduction and alkylation steps during sample preparation. The strict and relaxed target false discovery rates (FDR) were 0.01 and 0.05, respectively, input in the decoy database search parameters.

Results appear after running the workflow and opening the report. The number of proteins and peptides identified are shown. Proteins are listed with an accession number, a description, a score, a coverage, the number of unique peptides, the number of peptides, the number of peptide spectra matches, the number of amino acids, the molecular weight, and the calculated isolectric point. Proteome Discoverer assigns a protein score with the following equation:

$$Protein\ score = (sum\ of\ all\ cross - correlational\ factors\ of\ 0.8\ or\ above) + (peptide\ charge * peptide\ relevance\ factor)$$

The peptide relevance factor is a value between 0.0 and 0.8 specified by the user; the default is 0.4 (*40*). An XCorr value is achieved by scoring "the number of fragment ions that are common to two different peptides with the same precursor mass and [calculating] the cross-correlation score for all candidate peptides queried from the database" (*40*). The higher the Xcorr, the higher assignment confidence. Peptide matches can be viewed underneath their protein of origin as a drop-down. A peptide tab shows all the peptides identified. Important characteristics listed with peptides are sequence, number of peptide spectra matches, modifications, a delta Cn, an Xcorr, a charge, a molecular weight (MH+), a retention time, and the number of missed cleavages. The

delta Cn represents “the normalized score difference between the currently selected [peptide spectra match] and the highest scoring [peptide spectra match] for that spectrum” (*40*). A lower value is better for this score. Additionally, the program uses green, yellow, and red circles as confidence indicators (*40*). During the decoy database search, filters are applied to achieve the specified FDR, and the same filters are used to apply confidences (*40*). Green circles coincide with peptides that pass the strict FDR filter, yellow with those that pass the relaxed filter, and red with those that pass neither (*40*).

Chapter 3: Results and Discussion

3.1 Purified ArsR

A total of three purified ArsR samples were studied. Both the first and second sample were examined to evaluate digestion and sequence coverage. They also served as evidence that ArsR could be prepared and detected in isolation from the rest of the *H. pylori* proteome, confirming the proper functioning of the ESI source in the process. Additionally, since SEQUEST’s utility relies on the researcher’s ability to interpret scoring, the analysis of a purified protein sample of known concentration provided more support for its proper interpretation. The third sample was divided into two separate digestions and analyses to evaluate proteolysis with the Asp-N protease in combination with Trypsin.

The first sample was analyzed multiple times with LC-MS/MS using both NSI (Ch. 4) and ESI. Below is an example view of the SEQUEST report of one of the runs performed with ESI.

Proteins | Peptides | Search Input | Result Filters | Peptide Confidence | Search Summary

	Accession	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# AAs	MW [kDa]	calc. pI
1	15644795	response regulator ompR [Helicobacter pylori 26695]	68.36	57.33 %	1	13	23	52	225	25.8	5.43

	A2	Sequence	# PSMs	# Proteins	# Protein Groups	Protein Group Accessions	Modifications	ΔCn	XCorr	Charge	MH+ [Da]	ΔM [ppm]	RT [min]	# Missed Cleavages
1		KEEVSEPGDANIFR	3	1	1	15644795		0.0000	4.76	2	1592.62346	1160.07	14.57	1
2		ESIAIESESINPESSNK	2	1	1	15644795		0.0000	4.71	2	1834.53398	361.22	16.06	0
3		SHKKEEVSEPGDANIFR	3	1	1	15644795		0.0000	4.26	2	1943.64531	351.65	11.74	2
4		KEEVSEPGDANIFRVDK	2	1	1	15644795		0.0000	3.63	3	1935.90964	1520.42	15.21	2
5		DSREVYMHEK	1	1	1	15644795		0.0000	3.38	3	1296.27454	2071.54	3.18	1
6		NPKQPQYIISVR	1	1	1	15644795		0.0000	3.03	3	1447.00724	2899.62	14.45	1
7		EEVSEPGDANIFRVDK	1	1	1	15644795		0.0000	2.64	2	1806.40654	849.90	16.90	1
8		ALDYGADDYLPKPYDPK	4	1	1	15644795		0.0000	2.50	3	1942.81198	969.87	19.97	0
9		EEVSEPGDANIFR	2	1	1	15644795		0.0000	2.43	2	1463.33525	447.14	16.41	0
10		SDVEDKIK	1	1	1	15644795		0.0000	2.35	1	933.46313	-27.50	2.96	1
11		SIDVIIGR	3	1	1	15644795		0.0000	2.26	2	874.49217	2255.17	16.09	0
12		HIPIIISSAR	6	1	1	15644795		0.0000	2.03	2	1107.78447	1007.75	14.50	0
13		EVYMHEK	1	1	1	15644795		0.0000	1.83	1	935.46594	39.32	2.46	0
14		IQSLLR	2	1	1	15644795		0.0000	1.72	2	731.85039	3263.79	11.70	0
15		GYVFSR	1	1	1	15644795		0.0000	1.58	2	728.41100	52.68	11.24	0
16		KGYVFSR	4	1	1	15644795		0.0000	1.50	2	855.19512	-1487.94	8.33	1
17		SDVEDK	1	1	1	15644795		0.0000	1.43	1	692.58734	400.79	0.90	0
18		DSREVYmHEK	1	1	1	15644795	M7(Oxidation)	0.0000	1.42	3	1310.32578	565.97	2.27	1
19		QPQYIISVR	3	1	1	15644795		0.0000	1.40	2	1105.99956	2150.78	15.72	0
20		GIGYKLEY	3	1	1	15644795		0.0000	1.38	2	943.10503	648.81	17.78	1
21		hIPIIISSAR	1	1	1	15644795	N-Term(Acetyl)	0.1400	1.29	2	1147.21587	-1275.09	19.26	0
22		KLDLTR	1	1	1	15644795		0.0000	1.28	1	745.40521	-69.08	3.43	1
23		HIPIIISSARSDVEDK	1	1	1	15644795		0.0238	1.23	3	1780.88742	520.75	15.74	1
24		VDKDSR	2	1	1	15644795		0.0000	1.16	1	719.46313	131.84	0.81	1
25		VDKDSREVYMHEK	1	1	1	15644795		0.0806	1.14	3	1633.24582	-1551.38	3.15	2
26		kGYVFSR	1	1	1	15644795	N-Term(Acetyl)	0.0000	0.97	2	898.98479	563.56	11.05	1

Figure 3.1 Purified ArsR Sample 1, SEQUEST Report

ArsR is alternatively known as OmpR (gene HP0166) (*41*). SEQUEST identified its presence as “response regulator ompR” with a protein score of 68.36 and a sequence coverage of 57.33%. 26 peptides were matched to the ArsR sequence by SEQUEST; 13 represented high-confidence matches passing the strict FDR filter and 16 had Xcorr values greater than or equal to 1.50. KEEVSEPGDANIFR had 3 experimental peptide spectra matched to the theoretical spectrum generated for its sequence. The highest scoring match is displayed on the list above with an Xcorr of 4.76, and its experimental fragmentation spectrum can be viewed below.

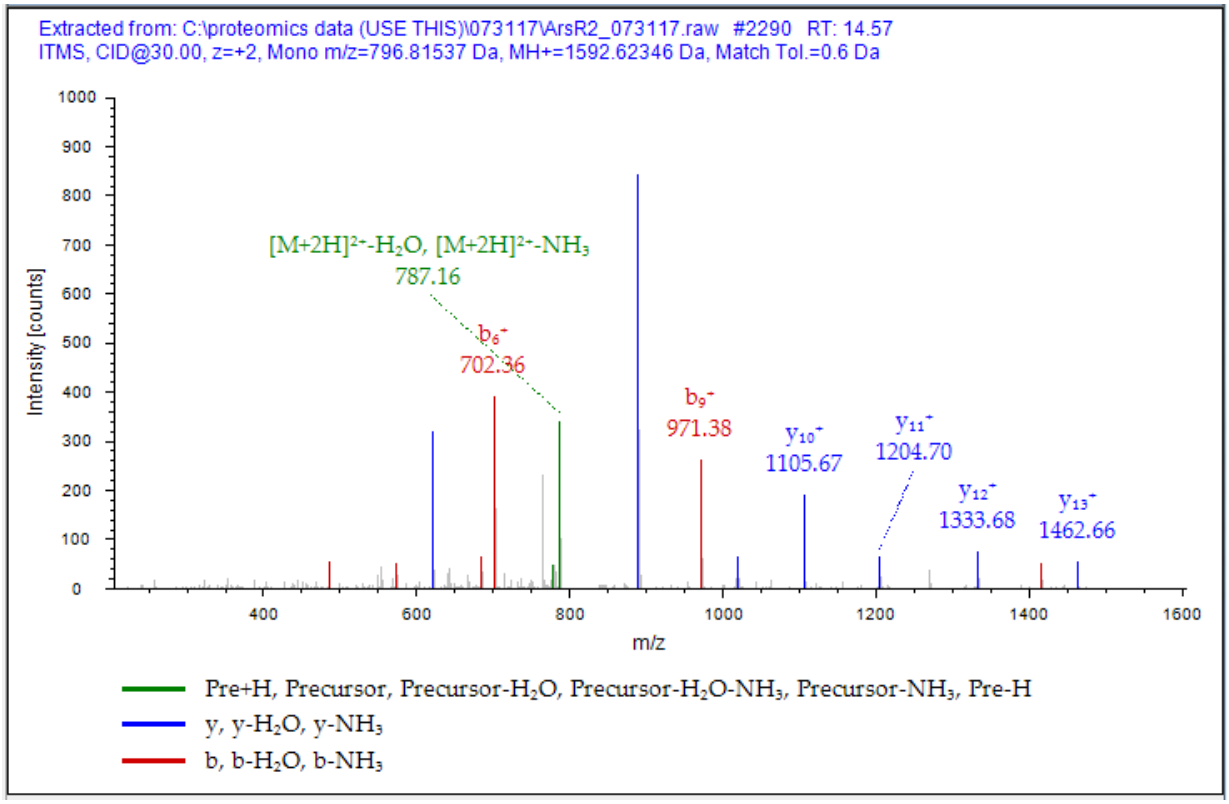

Figure 3.2 Product Ion Spectrum Matched to KEEVSEPGDANIFR Annotated by SEQUEST, Purified ArsR Sample 1

KEEVSEPGDANIFR was identified by a doubly-charged precursor peak of m/z 787.16 Da (loss of H_2O or NH_3); the monoisotopic m/z when doubly-charged is 796.82 Da. Its pseudomolecular ion m/z is 1592.62 Da. The peptide's retention time was 14.57 min. SEQUEST matched the CID fragment peaks (b and y ions) noted above to predicted fragments for the identified sequence, and then scored this spectrum against a theoretical spectrum.

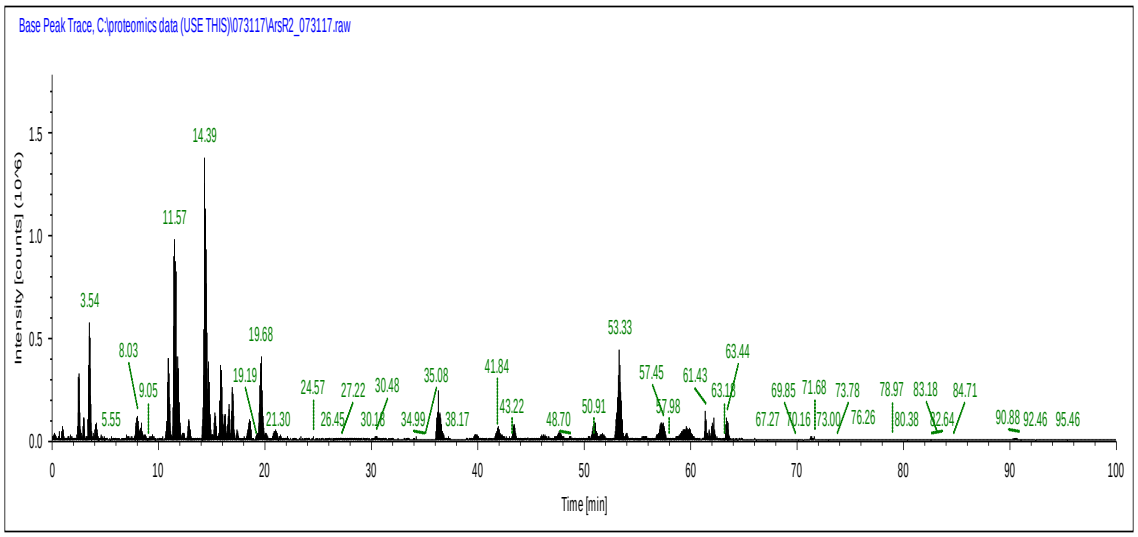

Figure 3.3 Chromatogram of Purified ArsR Sample 1

Noting the retention times reported in Fig 3.1, all the peptides identified for ArsR eluted before 20 minutes, when the percentage of organic solvent B was below 20%. Another view of this data can be seen below in the graph of peptide MH^+ vs retention time. In the chromatogram, the high intensity peaks grouped prior to 20 minutes were not matched to peptides in ArsR. The intensity counts for ions in the highest scoring experimental spectra matched to ArsR sequences were generally under 2,000 and never exceeded 8,500 except for the HIPIIISSAR sequence match which contained a y-ion that reached 400,000 (RT=14.50 min). The high intensity peaks prominently visible in the chromatogram may be contaminants because they were also seen in some *E. coli* samples run on the same analytical column. SEQUEST identified more than just peptides of the ArsR protein within the sample; these peaks can be attributed to proteins (digested into peptides) that escaped removal during Dr. Forsyth's ArsR purification.

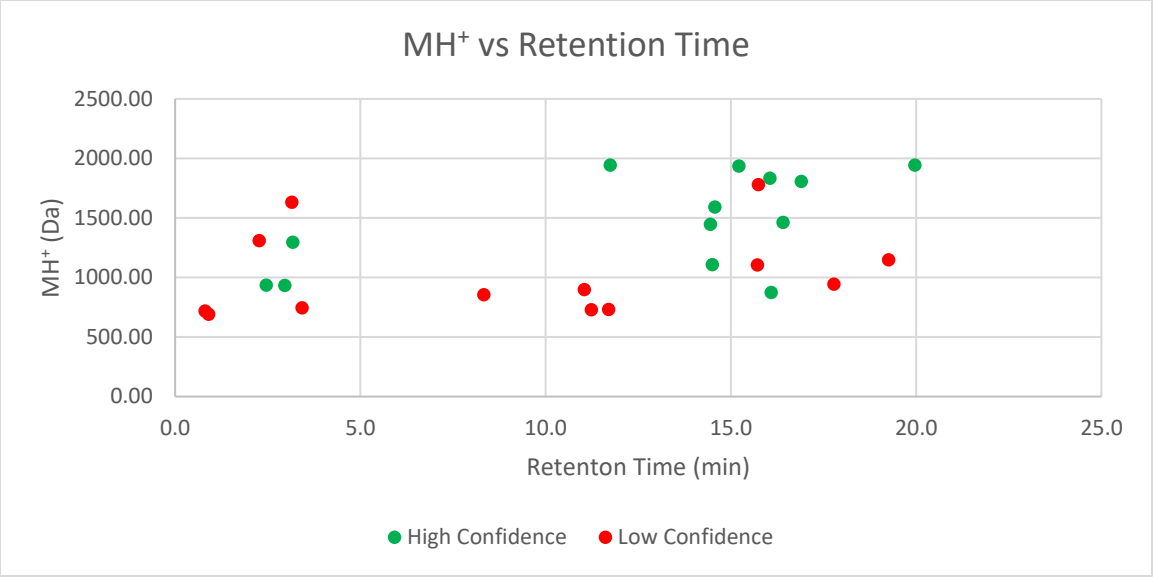

Figure 3.4 Peptide MH+ vs Retention Time, Purified ArsR Sample 1

As was mentioned above, this plot shows the retention times for ArsR sequence matches, all falling at or prior to 20 minutes in the HPLC gradient. This indicates that these peptides are decently hydrophilic, eluting before the percent organic solvent rose above 20.

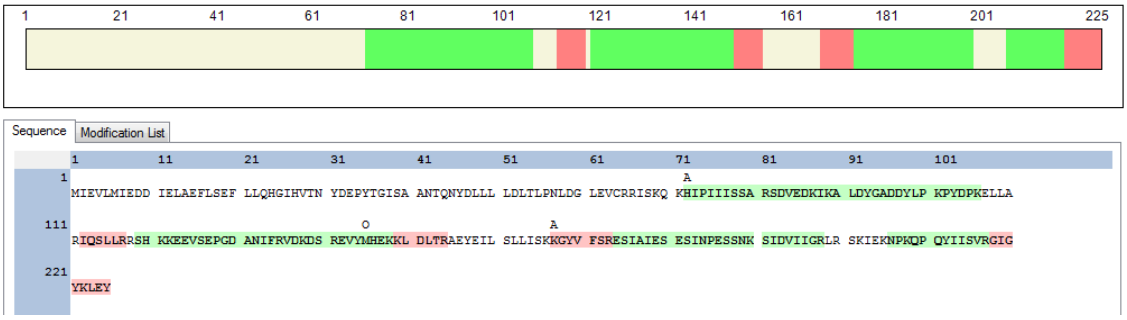

Figure 3.5 Sequence Coverage of Purified ArsR Sample 1

After the confident identification of peptides of ArsR, sequence coverage could be evaluated. The image above shows the sequence of ArsR (without histidine tag). The portions of the sequence represented by peptides are highlighted with their confidence indicator color. The aspartic acid residues of interest, D47, D52, and D59, were not represented at all in this sample. Upon closer inspection, it was realized that there were neither lysines nor arginines present in the

first 65 residues. Since those are the sites of action for trypsin's digestion, and the 65-residue length exceeded the mass range of the LTQ, a new enzyme, Asp-N, was considered. A second purified ArsR sample confirmed these findings.

Table 3.1 Purified ArsR Sample 2, High Confidence Peptide Matches

Sequence	MH+ (Da)	RT (min)	Xcorr	Charge
KEEVSEPGDANIFR	1590.04	15.11	5.11	3
SHKKEEVSEPGDANIFR	1943.27	12.35	4.43	2
ESIAIESESINPESSNK	1831.10	16.81	3.52	3
AEYEILSLLISK	1376.47	28.86	3.29	2
ALDYGADDYLPKPYDPK	1941.77	20.61	3.21	3
AEYEILSLLISKK	1507.10	26.27	3.06	2
KEEVSEPGDANIFRVDK	1932.52	15.54	2.86	3
HIPIIISSAR	1107.44	15.39	2.83	3
SDVEDKIK	933.84	3.49	2.81	2
KGYVFSR	855.24	9.38	2.58	2
SIDVIIGR	872.43	16.80	2.47	2
EEVSEPGDANIFR	1463.17	17.64	2.28	2
KLDLTR	747.90	5.08	2.20	2
GIGYKLEY	944.82	17.83	2.12	2
IQSLLR	729.45	12.47	1.50	1
SDVEDK	692.40	0.70	1.49	1

SEQUEST identified ArsR as response regulator ompR with a protein score of 103.15 and sequence coverage of 64.89%. 33 peptides were matched to the protein sequence; 16 were matched with high confidence, and 21 had Xcorr values above 1.50. KEEVSEPGDANIFR was, again, the peptide sequence matched with the highest Xcorr (5.11); in this sample run, it had 9 peptide spectral matches. The highest scoring experimental spectrum match is seen below.

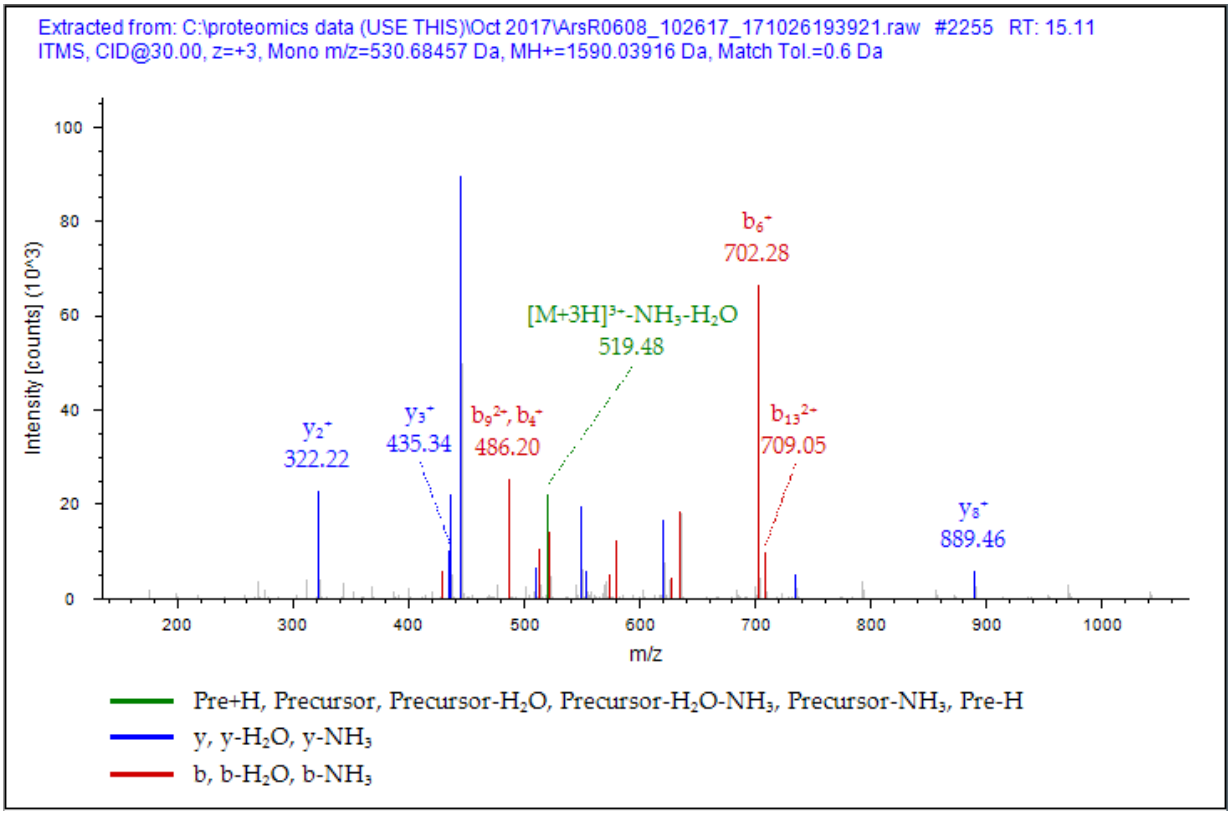

Figure 3.6 Product Ion Spectrum Matched to KEEVSEPGDANIFR Annotated by SEQUEST, Purified ArsR Sample 2

The MH^+ of KEEVSEPGDANIFR was reported as 1590.04 Da, and its recognized precursor ion was triply-charged, with a loss of NH_3 or H_2O. Its monoisoptopic m/z was 530.68 Da. Its retention time was 15.11. Not only is there confidence in its assignment based on Xcorr and false discovery rates, but the consistency in SEQUEST matching this sequence in the first sample and this sequence in the second is favorable. The retention times are only different by 0.54 minutes or about 32 seconds. Its listed precursor mass falls within the user-specified tolerance of ±1.6 Da. Identified ions' intensities fell between thousands of counts and 100,0000 counts, providing support for their assignments as peptide fragments, in contrast to the counts under 2,000 observed for many experimental peptide product-ion spectra.

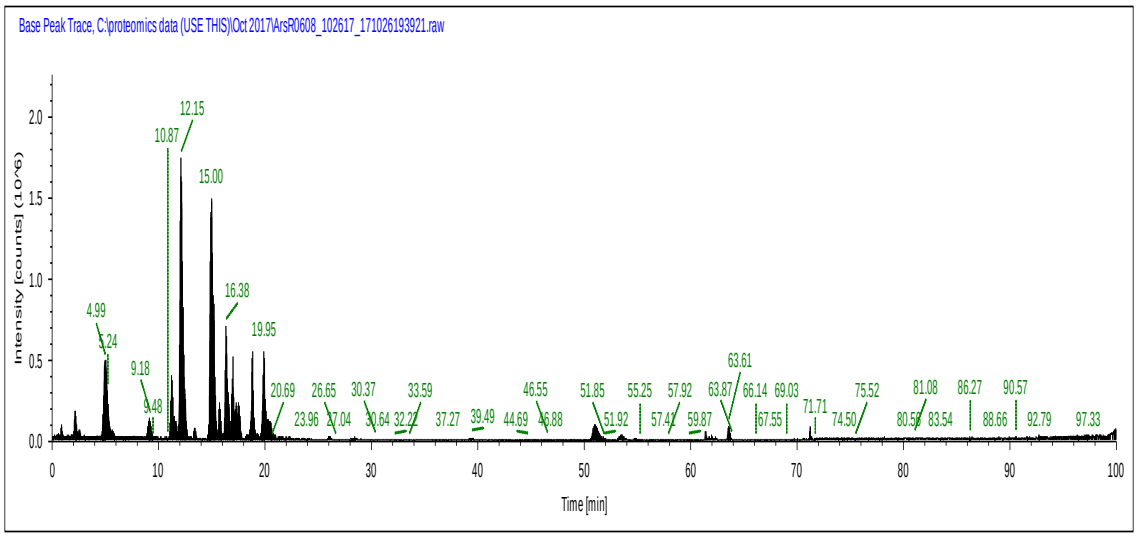

Figure 3.7 Chromatogram of Purified ArsR Sample 2

Comparing the chromatograms of sample 1 and sample 2, there is a similar grouping of high intensity peaks before about 20 minutes. These, for example the 12.15 minute and 15.00 minute peaks identified on the spectrum, were not matched to peptides. Apart from a few, the assignments for peptides from the ArsR sequence again eluted in this region. Consistency and high-confidence peptide matches indicated the analytical column's integrity and supported SEQUEST's interpretations of the data. Additionally, if a sample with a more complex mixture of peptides were analyzed under these same conditions, retention times from about 10 to 20 minutes could help evaluate SEQUEST's matches for ArsR-derived peptides, at least for peptide fragments past the 65th residue.

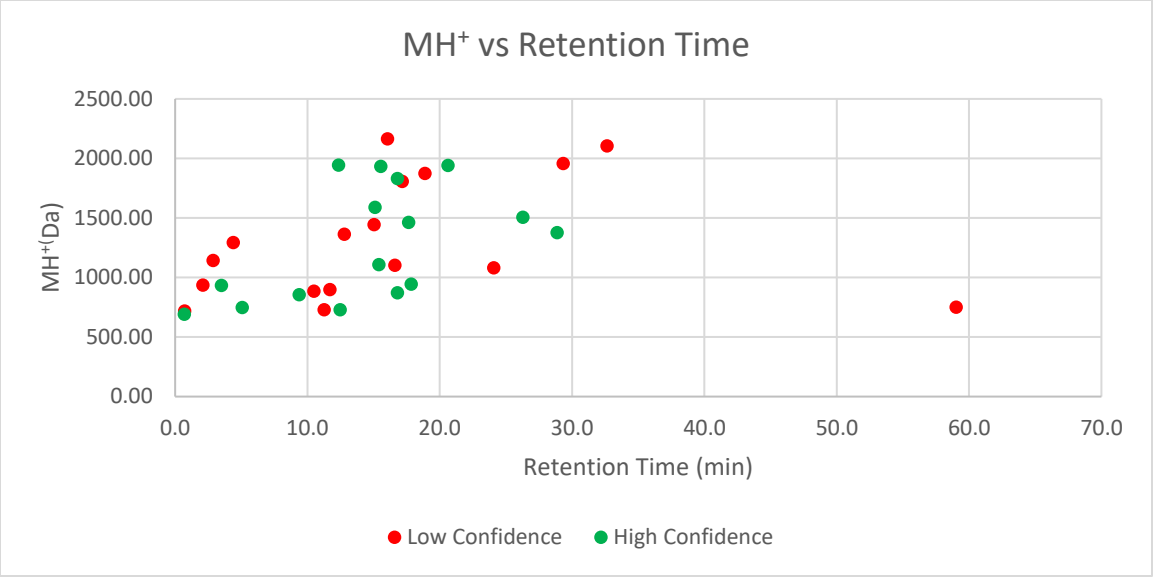

Figure 3.8 Peptide MH^+ vs Retention Time, Purified ArsR Sample 2

Above is another representation displaying the abundance of elution prior to 20 minutes. In this sample, there were some high confidence matches slightly outside this region in the 20 to 30-minute range.

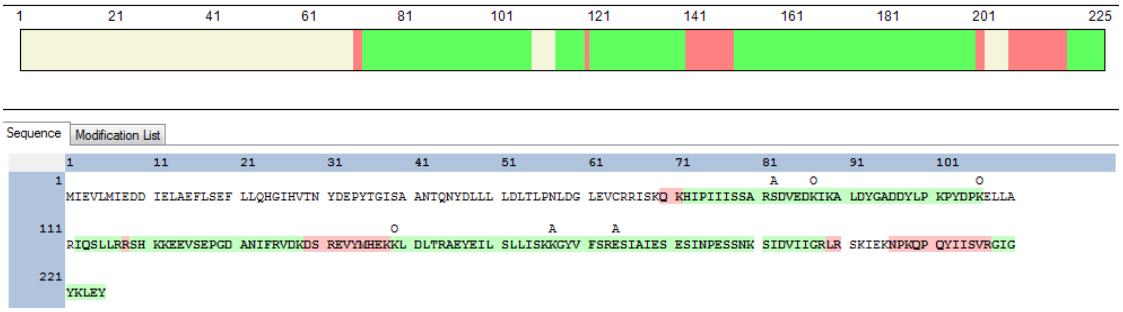

Figure 3.9 Sequence Coverage of Purified ArsR Sample 2

The sequence of ArsR covered in sample 2 contains nearly everything past lysine residue 69. Again, the aspartic acid residues to be analyzed for phosphorylation were not present, indicating the need for another enzyme to break up the beginning of the protein. Asp-N was chosen for testing.

Table 3.2 Purified ArsR Asp-N and Minimal Trypsin Digestion, Asp-N Work-up

Sequence	MH+ (Da)	RT (min)	Xcorr	Charge
DEPYTGISAANTQNY	1644.16	17.74	3.04	2
DANIFRV	833.43	18.35	2.18	2
DDYLPKPY	1010.46	16.74	1.43	1
DLTLPNL	785.21	23.65	1.25	1
DYLPKPY	895.60	16.97	1.13	2
DKIKALDYGAd	1225.79	26.02	0.67	2
dANIFRV	875.12	21.76	0.62	2
MIEVLMIE	978.00	0.85	0.55	2
DKIKAl	703.87	71.28	0.54	2
mIEVLMIEDDIELAEFLSEFLLQHGIHVTNy	3693.20	47.02	0.40	3
DKIKALDYGAD	1209.79	42.53	0.34	1
DYGADDYLPKPy	1433.95	47.35	0.24	1

Since SEQUEST can only use certain combinations of enzymes for its analysis, the data for digestions utilizing trypsin and Asp-N were analyzed three separate times: with Asp-N selected, trypsin selected, and no enzyme selected. The purified ArsR sample for the first evaluation of Asp-N was prepared with a lower amount of trypsin than used for the previous two samples. For the first run of this sample, SEQUEST produced the report above when Asp-N was selected as the cleavage enzyme. 12 peptide sequences were matched to the data; 4 were matched with high confidence, and 2 had Xcorr values over 1.50. SEQUEST's protein score for ArsR was 11.20 with a sequence coverage of 34.67%. The highest scoring peptide match had an Xcorr of 3.04. The chromatogram for this run can be viewed below.

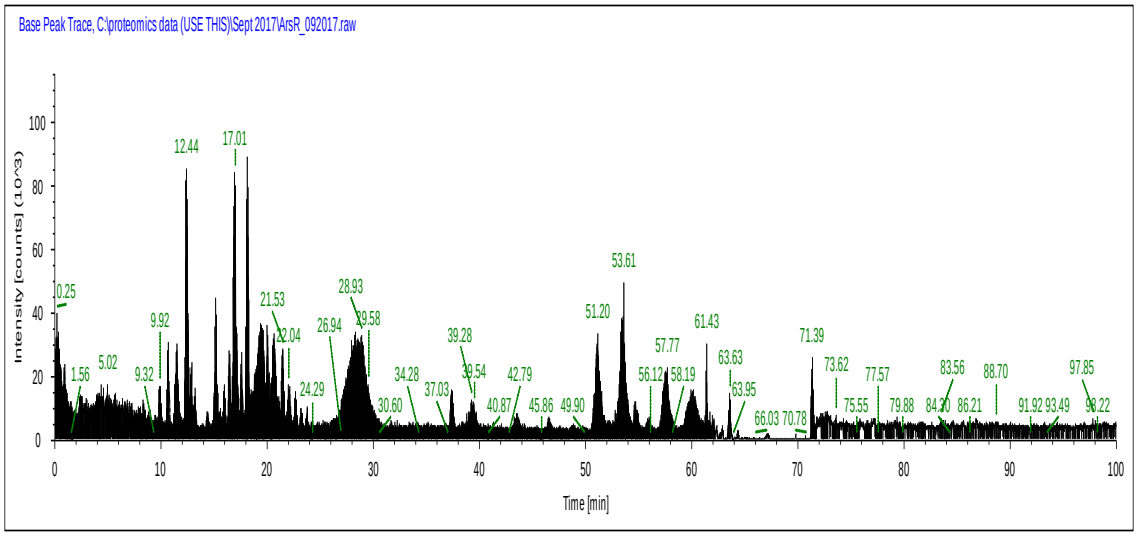

Figure 3.10 Chromatogram of Purified ArsR Asp-N and Minimal Trypsin Digestion

While the region between 10 and 20 minutes seems to contain some similar high intensity peaks to the trypsin and Lys-C digestions, this chromatogram shows more species eluting from the analytical column overall. To see where the Asp-N cleaved peptides eluted, the following plot can be consulted.

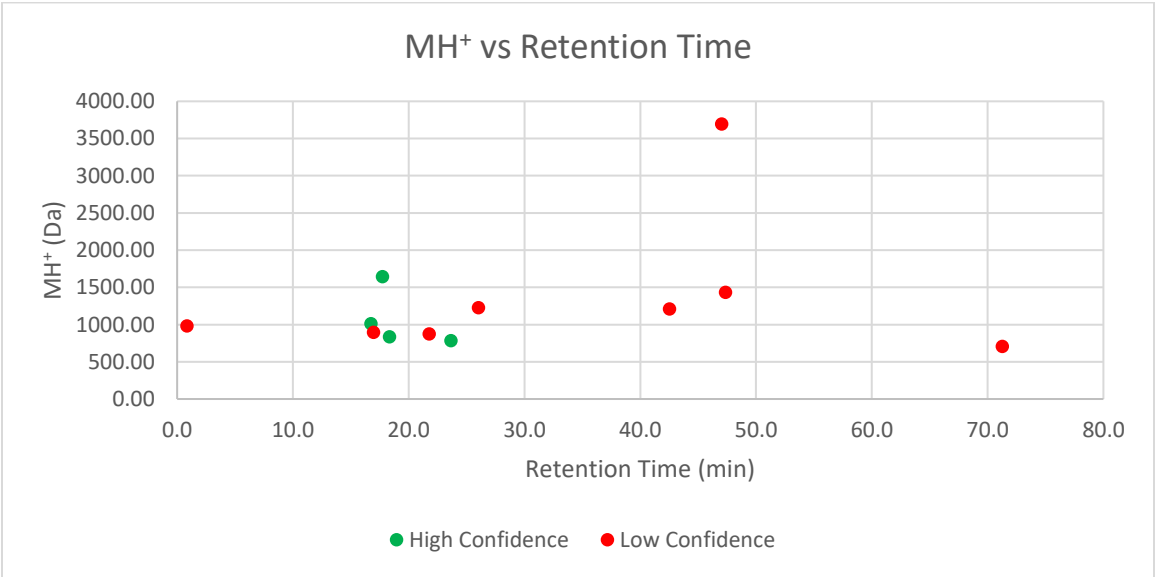

Figure 3.11 Peptide MH^+ vs Retention Time, Purified ArsR Asp-N and Minimal Trypsin Digestion, Asp-N Work-up

While there is still a grouping present towards the beginning of the gradient, and the high confidence peptide sequence matches eluted in that range, there exists a greater spread of elution recognized by SEQUEST than in the trypsin and Lys-C digestion.

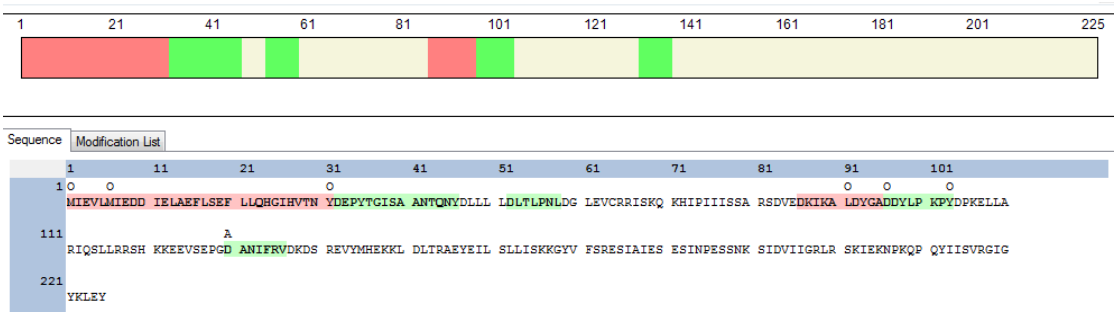

Figure 3.12 Sequence Coverage of Purified ArsR Asp-N and Minimal Trypsin Digestion, Asp-N Work-up

The Asp-N workup showed improved coverage of the beginning of the ArsR protein sequence. Notably, aspartic acid residue 52 was present and matched with high confidence. The precursor ion matched to the D52 containing sequence (DLTLPNL) was only singly-charged. This may be problematic for future application of ETD for PTM analysis. The other two residues, D47 and D59, were not observed with Asp-N selected in SEQUEST.

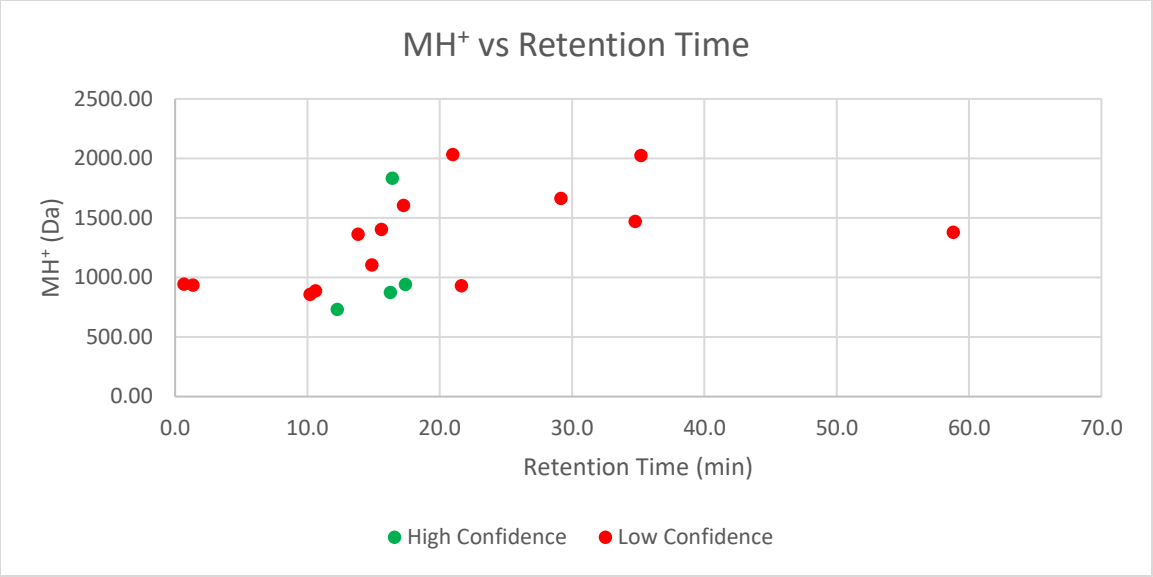

Figure 3.13 Peptide MH^+ vs Retention Time, Purified ArsR Asp-N and Minimal Trypsin Digestion, Trypsin Work-up

The trypsin work-up still indicated some grouping of peptides eluting before 20 minutes, but it also displayed some lower confidence peptides eluting a little later in the gradient at higher organic solvent concentrations.

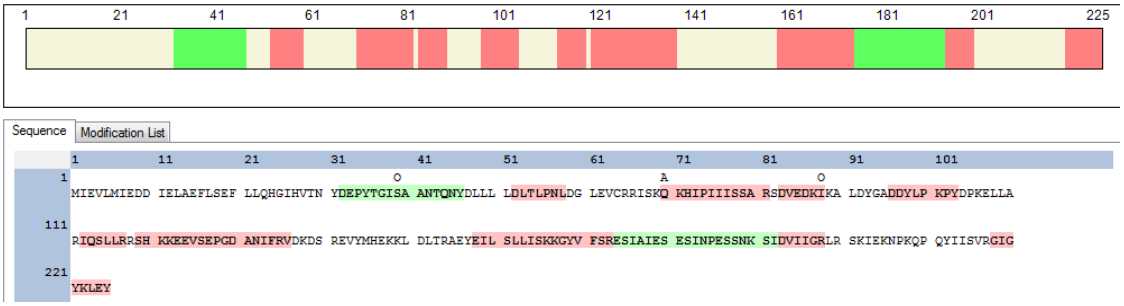

Figure 3.14 Sequence Coverage of Purified ArsR Asp-N and Minimal Trypsin Digestion, No Enzyme Work-up

With no enzyme selected, SEQUEST identified response regulator ompR with a protein score of 25.72 and a sequence coverage of 53.79%. Only D52 was present, not D47 or D59. In this workup, the precursor ion matched to DLTLPNL was still singly-charged, but was matched with lower confidence. Portions of the end of the protein seen in previous digestions were not present

due to the lower amount of trypsin enzyme used in sample preparation. Before moving onto another digestion, the sample was run a second time to confirm these results. Additionally, during the spin column procedure, flow-throughs were retained in case of inefficient sample binding. The procedure notes that if the sample is not sufficiently hydrophobic, it may not bind well to the resin bed during de-salting. Noting the peptide elutions observed at the beginning of the gradient in mind, these flow-throughs were analyzed. The flow-through retained from the first preparation with minimal trypsin was filtered again through a PVDF syringe filter in attempts to clean-up the flow-through a bit; it seemed it would be difficult to identify peptides remaining in the solution with the salts competing for ionization.

The second analysis of the Asp-N and minimal trypsin digested sample yielded similar results to the first, again, providing confidence in the setup, and permitting evaluation of sample preparation.

Table 3.3 Purified ArsR Asp-N and Minimal Trypsin Digestion, LC-MS/MS Run 2, Asp-N Work-up

Sequence	MH+ [Da]	RT [min]	XCorr	Charge
DEPYTGISAANTQNY	1643.89	17.71	2.75	2
DDYLPKPY	1010.48	16.62	1.94	1
DLTLPNL	785.28	23.58	1.57	1
DANIFRV	837.21	17.62	1.32	2
DYLPKPY	895.39	16.48	0.97	1
MIEVLmIEd	1123.87	16.70	0.94	2
DSREVYMHEKKL	1532.06	34.93	0.83	3
dANIFRV	876.42	27.74	0.68	2
mIEVLmIE	1010.32	24.86	0.60	2
dANIFRVDK	1119.76	46.58	0.38	1
dKIKALDYGAD	1251.03	30.84	0.21	1

The Asp-N work-up of the second LC-MS/MS analysis yielded the identification of response regulator ompR with a protein score of 11.27 and a sequence coverage of 31.11%. 11

peptides were matched to the ArsR sequence, with 3 of a high confidence and 3 with Xcorr's of greater than 1.50. The highest Xcorr was 2.75 for the sequence, DEPYTGISAANTQNY. This sequence was the highest scoring for the first run as well (Table 3.2).

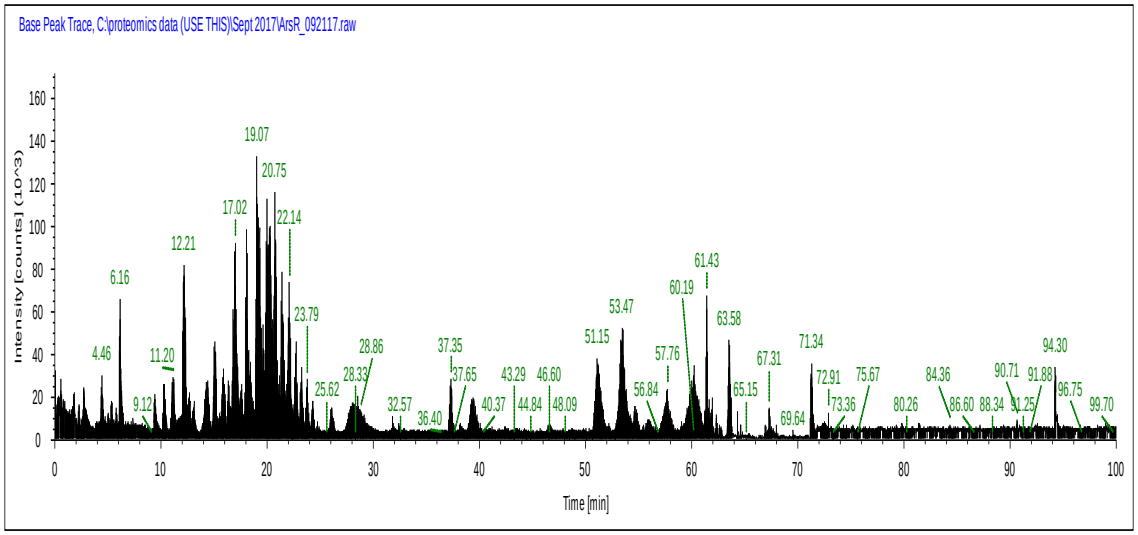

Figure 3.15 Chromatogram of Purified ArsR Asp-N and Minimal Trypsin Digestion, LC-MS/MS Run 2

The grouping of chromatogram peaks around 20 minutes was very similar to the last analysis. For example, peaks around 17 and 18 are present in both chromatograms. There are identified peaks at 16.62 (DDYLPKPY), 16.70 (MIEVLmIEd), 17.62 (DANIFRV), and 17.71 minutes (DEPYTGISAANTQNY) in this sample. In the previous analysis, there were identified peaks at 16.74 (DDYLPKPY), 16.97 (DYLPKPY), 17.74 (DEPYTGISAANTQNY), and 18.35 minutes (DANIFRV). There are many higher intensity peaks present in this chromatogram compared to the trypsin/Lys-C-only digest, consistent with the first run of this sample.

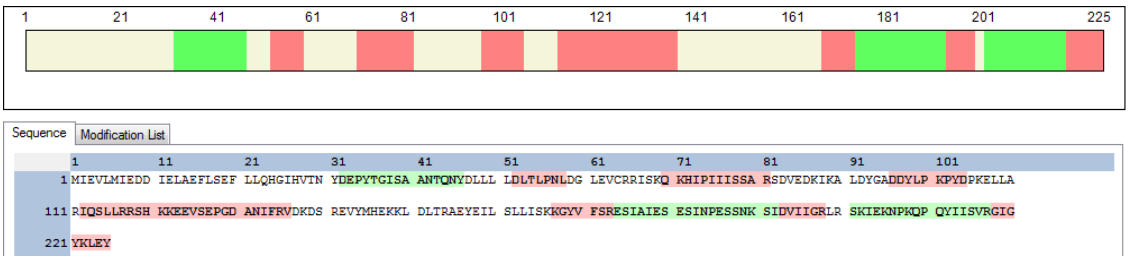

Figure 3.16 Sequence Coverage of Purified ArsR Asp-N and Minimal Trypsin Digestion, No Enzyme Work-up, LC-MS/MS Run 2

The no enzyme work-up yielded an ArsR protein score of 35.51 and a sequence coverage of 55.56%. These were around the same values as the last run. Aspartic acid residue D52 was still the only residue of interest represented. The precursor ion matched to DLTLPNL was doubly-charged for this no-enzyme workup, and was identified with low confidence.

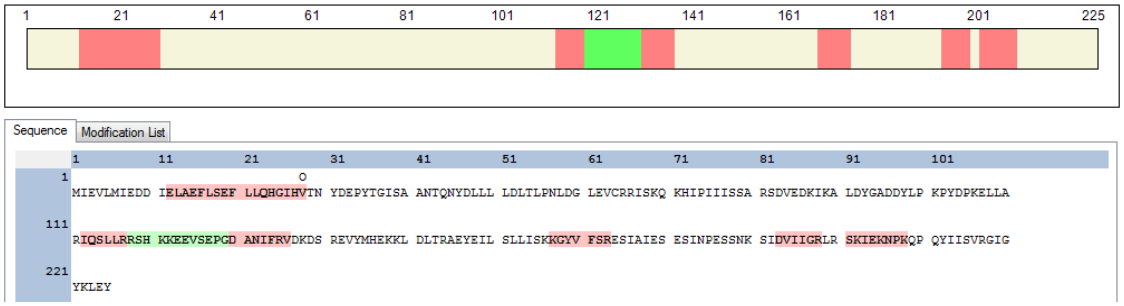

Figure 3.17 Sequence Coverage of Spin Column Flow-Through of Purified ArsR Asp-N and Minimal Trypsin Digestion, No Enzyme Work-up

At first, when the LC-MS/MS run of this flow-through was worked up with Asp-N or trypsin selected, ArsR was given a protein score of 0.00, so the presence of sample remaining in the flow-through seemed to be minimal; however, when the same spectra were analyzed without specifying the enzyme, the protein scored 5.07 and had a sequence coverage of 28.00%. This is different from zero, so there must have been peptides that did not bind well to the resin bed; they may have been insufficiently hydrophobic. Their distribution by retention time can be viewed below.

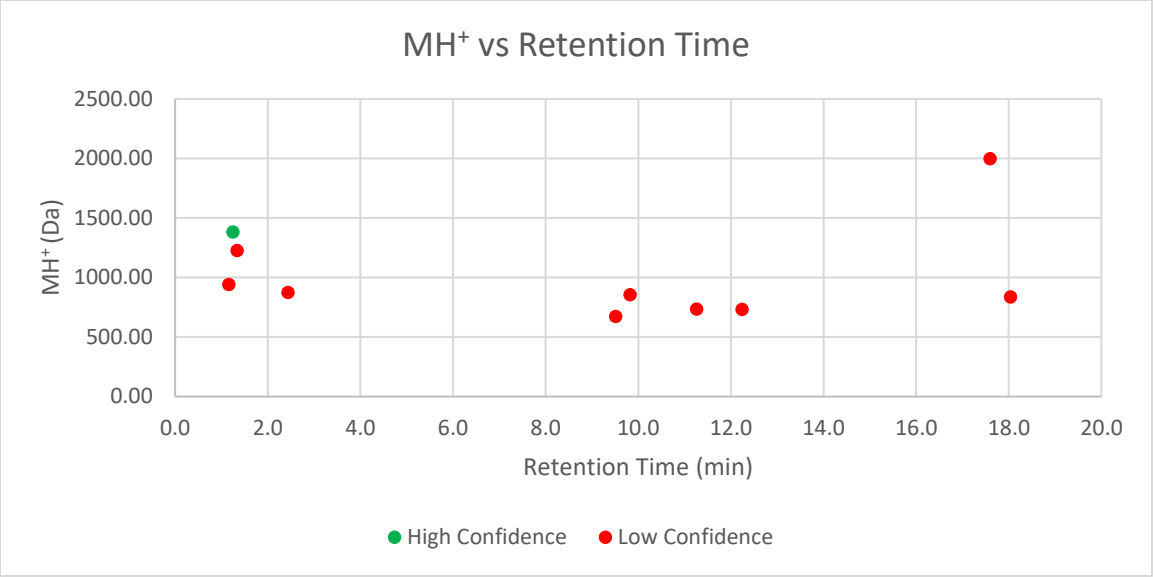

Figure 3.18 Peptide MH^+ vs Retention Time, Spin Column Flow-Through of Purified ArsR Asp-N and Minimal Trypsin Digestion, No Enzyme Work-up

Each of the peptides matched to the ArsR sequence eluted prior to 20 minutes, with 4 peptides, including one matched with high confidence, eluting before 4 minutes. This indicates that there are peptides in ArsR that are polar enough to hinder binding to the C18 resin bed during sample clean-up.

The next sample preparation involved the use of more trypsin at a similar substrate-to-enzyme ratio to the trypsin and Lys-C purified ArsR digestions.

Table 3.4 Purified ArsR Asp-N/Trypsin Digestion, Asp-N Work-up

Sequence	MH+ [Da]	RT [min]	XCorr	Charge
DEPYTGISAANTQNY	1643.82	17.52	2.38	2
DANIFRV	833.04	17.95	2.05	2
DLLLLDLTLPNL	1352.77	36.98	1.94	2
DYGADDYLPKPY	1418.99	19.38	1.88	2
DDYLPKPY	1010.92	16.65	1.78	2
DLTLPNL	785.29	23.41	1.50	1
DYLPKPY	895.17	16.78	1.23	2
MIEVLMIE	979.46	16.21	0.78	2
MIEVLmIEd	1124.11	17.65	0.73	2
DKIKAL	686.30	19.72	0.66	1
dANIFRVDk	1134.29	19.84	0.64	2
DKIKALDYGA	1092.29	24.94	0.64	2
mIEVLMIED	1107.21	19.49	0.61	2
MIEVLmIE	993.01	24.38	0.46	2
dYLPKPY	936.48	72.85	0.46	2
DKIKALDYGAD	1209.17	43.15	0.29	1

For the Asp-N workup, SEQUEST identified the above sequences from ArsR with a protein score of 15.59 and a sequence coverage of 28.00%. 16 peptide sequences were matched to the experimental spectra. 1 high confidence sequence was reported, and 6 had Xcorr scores of greater than or equal to 1.50. This sample had the same highest scoring peptide in common with both LC-MS/MS runs for the previous digestion when worked up in this manner.

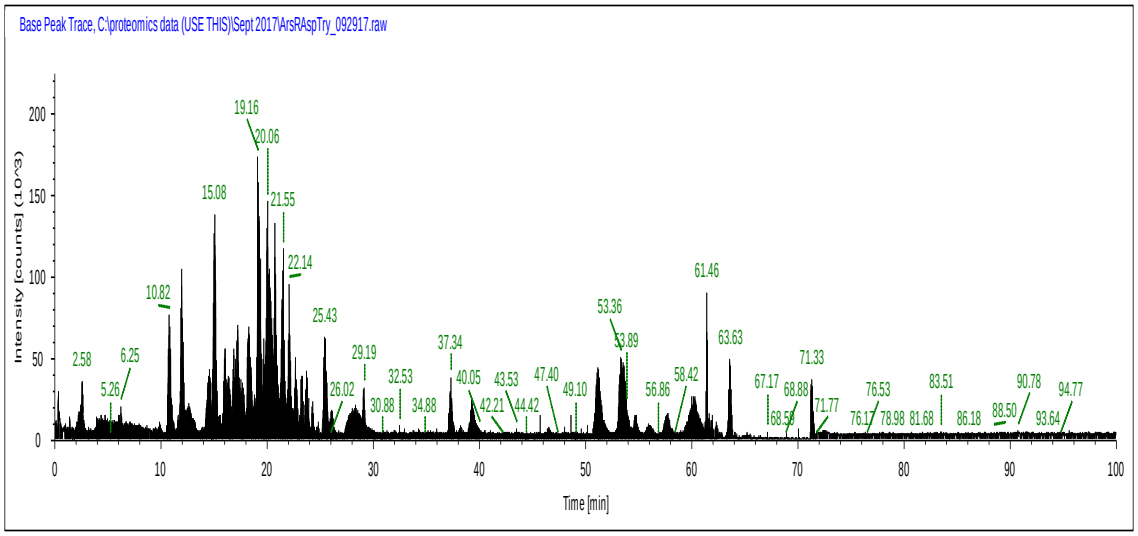

Figure 3.19 Chromatogram of Purified ArsR Asp-N/Trypsin Digestion

The chromatogram takes on similar characteristics to the first Asp-N digestion's chromatogram. There is a grouping of peptide elution around the 10 to 20-minute range. For example, DEPYTGISAANTQNY was identified again with a retention time of 17.52 minutes. Ion counts for fragment ions matched to sequences in ArsR were below 1,000. The high intensity peaks in this elution range (like RT 19.16 min and 20.06 min) were not matched to peptides.

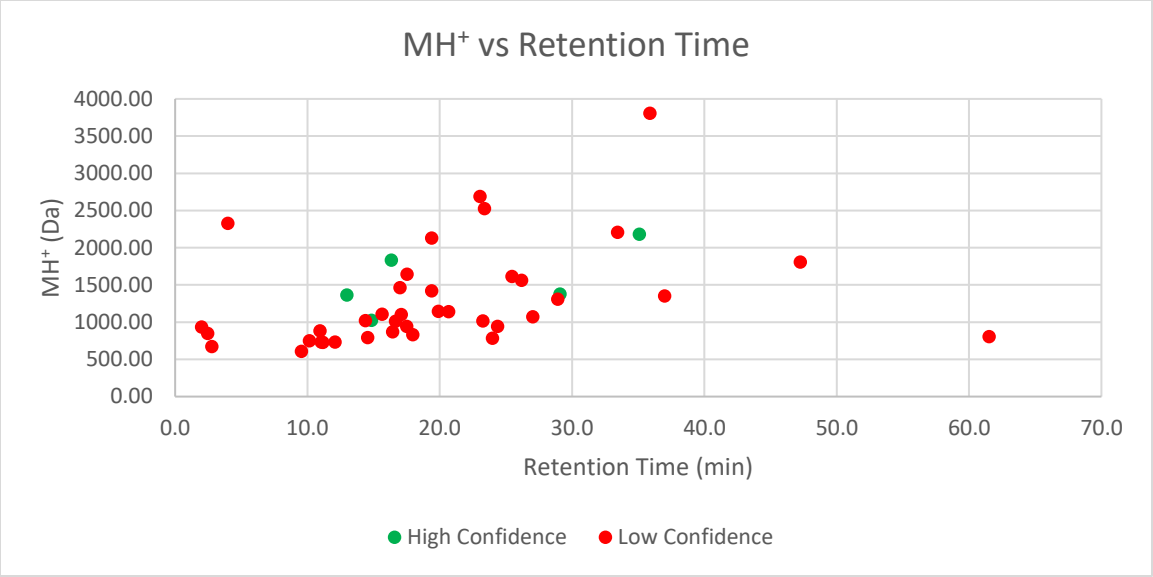

Figure 3.20 Peptide MH+ vs Retention Time, Purified ArsR Asp-N/Trypsin Digestion, No Enzyme Work-up

Above is the distribution of matched peptides and their retention times for the no enzyme work-up. In agreement with the previous samples, peptides identified as deriving from ArsR eluted mainly towards the beginning of the gradient.

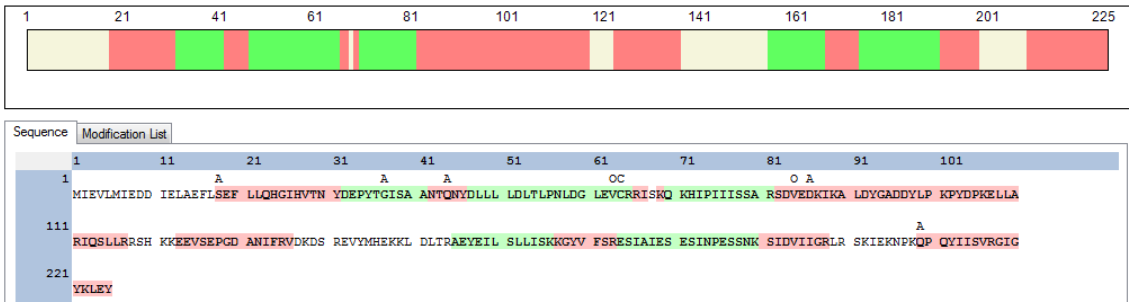

Figure 3.21 Sequence Coverage of Purified ArsR Asp-N/Trypsin Digestion, No Enzyme Work-up

For the no enzyme work-up, SEQUEST scored ArsR, again, always recognized as "response regulator ompR" in the *H. pylori* FASTA, as 67.05 with a sequence coverage of 77.33%. Importantly, D47, D52, and D59 were identified within the matched sequences, and these sequences were matched with high confidence. It should be noted that only D47 and D52 were identified when Asp-N was input as the cleavage enzyme (Fig 3.22). Additionally, D47 was within

a low confidence peptide. To try to gain more confidence in the D47 residue's peptide match, the minimum peptide length was decreased from 6 to 5 in the SEQUEST parameters and analyzed again with Asp-N as the cleavage enzyme (Fig 3.23). In the sequence of ArsR, D47 is followed by four leucine residues and then another aspartic acid. Since Asp-N cleaves on the N-terminal side of aspartic acid residues, it was possible that there were peptides of 5 residue-length not being noticed by SEQUEST. While this did not significantly improve coverage, inputting no enzyme parameter did. As mentioned earlier, SEQUEST cannot look for both trypsin and Asp-N cleavages simultaneously. A peptide identified with high confidence without specifying an enzyme was between aspartic acid 47 and arginine 65 (DLLLLDLTLPNLDGLEVcR, carbidomethylation of cysteine present, Xcorr: 3.25). Asp-N does not cleave on the carboxy-terminal side of arginine residues; it was not identified by SEQUEST when Asp-N was input as the proteolytic enzyme. The sequence contained D47, D52, and D59 and was doubly-charged, granting the possibility for ETD and the analysis of combinatorial PTMs.

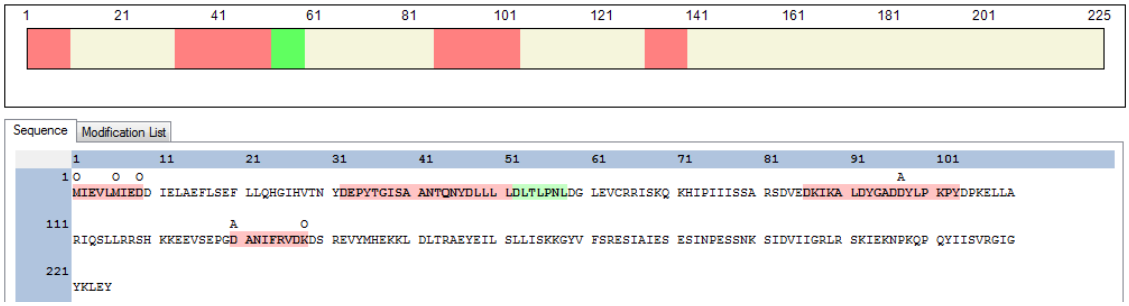

Figure 3.22 Sequence Coverage of Purified ArsR Asp-N/Trypsin Digestion, Asp-N Work-up

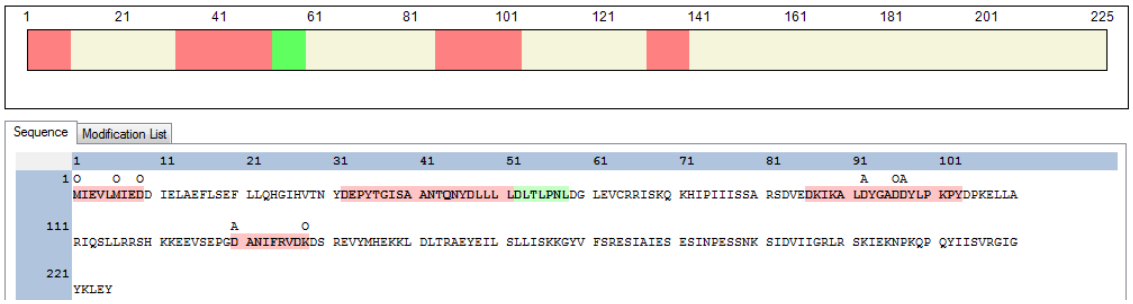

Figure 3.23 Sequence Coverage of Purified ArsR Asp-N/Trypsin Digestion, Asp-N work-up, Minimum Peptide Length of 5

As can be seen, figures 3.22 and 3.23 look the same. Decreasing the minimum peptide length identified a peptide sequence DLLLL, already present in a larger, low confidence sequence. This new identification did not increase confidence, nor did it increase sequence coverage.

The flow-through from this sample's preparation was also retained and tested, this time without filtering by PVDF syringe filter prior to LC-MS/MS.

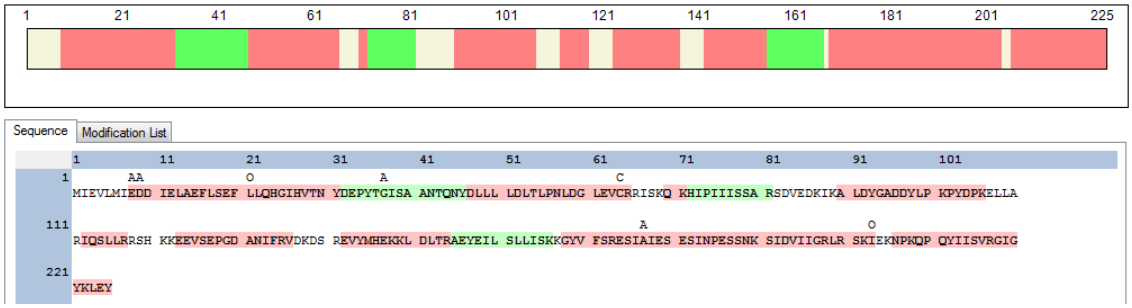

Figure 3.24 Sequence Coverage of Spin Column Flow-Through of Purified ArsR Asp-N/Trypsin Digestion, No-Enzyme Work-up

SEQUEST reported ArsR with a protein score of 58.93 and a sequence coverage of 83.56%. This is incredibly notable for a flow-through that wasn't supposed to contain any sample. Additionally, D47, D52, and D59 were identified in peptides matched to the experimental spectra. The precursor ions containing the residues-of-interest were doubly- or triply-charged.

3.2 Shotgun Proteomics

During the shotgun sample preparation, a BCA assay was performed to determine the protein concentration in solution. Cu^{1+} ions were produced from the reduction of Cu^{2+} by the protein's backbone, and these ions reacted with bicinchoninic acid (BCA), forming a purple colored product that absorbs at 562nm (*42*). A microplate UV/Vis and its accompanying software were used measure absorbances at 562nm and to create a calibration curve, relying on Beer's law. Beer's law expresses that absorption is directly proportional to the concentration of the solution and directly proportional to the path length of the light. A blank of ultrapure H_2O was measured and subtracted from the absorbance readings of the standards and sample. The absorbance data and calibration curve can be viewed below.

Table 3.5 BCA Assay Absorbances and Concentrations

	Blank	A	B	C	D	E	F	G	H	Sample
Absorbance	0.175	2.768	2.074	1.525	1.36	0.922	0.564	0.406	0.281	0.957
Absorbance Minus Blank	0	2.593	1.899	1.35	1.185	0.747	0.389	0.231	0.106	0.782
Concentration (µg/ml)	<0.000	2034.595	1432.104	973.987	840.02	494.574	226.799	115.254	32.576	521.556

A curve was prepared by plotting the blank-corrected absorbance readings vs the standard concentrations and fitting the data. The BCA protocol recommended a quadratic curve be used with the curve-fitting algorithms of the microplate reader, accounting for instrumental deviation from the conditions required for Beer's law.

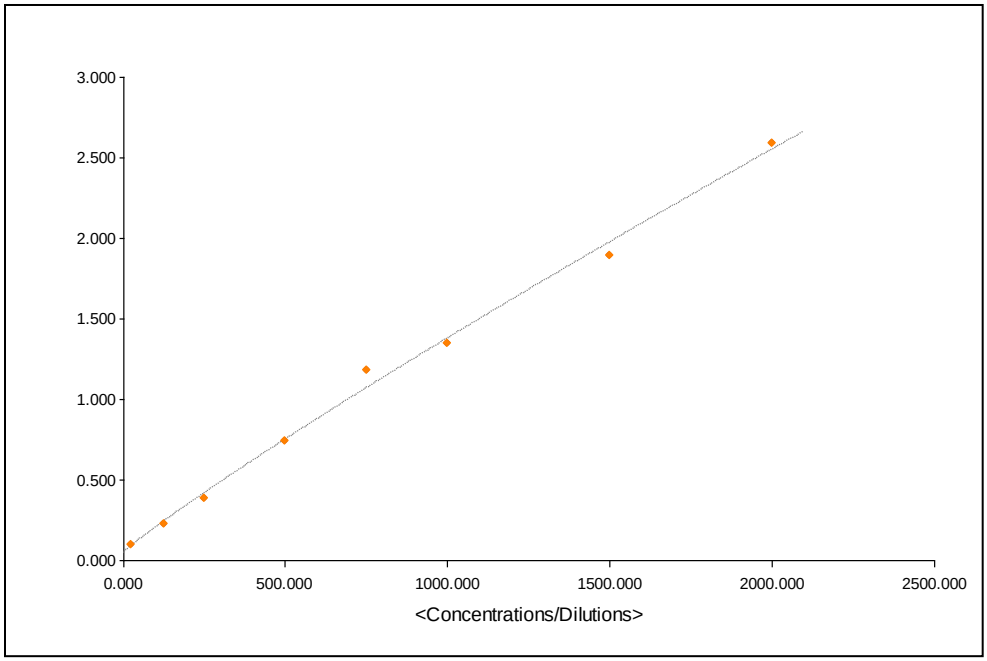

Figure 3.25 Standard Curve Produced by Gen5 Microplate Reader Program

The equation for the best quadratic fit can be viewed in the table below.

Table 3.6 BCA Assay Standard Curve Equation

Curve Name	Curve Formula	A	B	C	D	R2
StdCurve	Y = (A-D)/(1+(X/C)^B) + D	0.05	0.93	9.40E+04	92.3	0.996

Once the protein concentration of the sample was determined to be about 0.5 mg/ml, the shotgun sample was prepared with the proper amounts of reagents. Once prepared, the shotgun sample was run through LC-MS/MS three times.

Table 3.7 Shotgun Sample Analysis, 10/12

Sequence	MH+ [Da]	RT [min]	XCorr	Charge
KEEVSEPGDANIFR	1593.38706	15.12	1.83	3
ESIAIESESINPESSNK	1834.45146	16.50	1.77	2
AEYEILSLLISK	1376.12383	28.98	1.25	2
eLLARIQSLLRR	1509.85782	17.40	1.06	3
IQSLLRr	902.89897	23.95	0.72	2
aLDYGADDYLPKPYDPKELLAr	2580.04399	60.29	0.23	2

The first analysis was performed on October 12, 2017. SEQUEST indicated an ArsR protein score of 1.77 and a percent coverage of 32.00%. The 6 peptides matched from the ArsR sequence were of low confidence. The sample was run again since the confidence values were so low. It could not be assumed that these assignments would be consistent or allow any post-translational modification analysis on subsequent samples.

Table 3.8 Shotgun Sample Analysis, 10/20

Sequence	MH+ [Da]	RT [min]	XCorr	Charge
IEKNPKQPQYIISVR	1813.62558	22.58	1.25	3
SIDVIIGR	870.65972	16.10	1.18	2
HIPIIISSAr	1121.50737	14.29	1.16	2
QKHIPIIISSAR	1360.24637	18.53	1.10	3
lRSKIEK	914.08629	9.59	0.81	2
VDKDSr	735.65929	13.22	0.70	2
gIGYKLEY	983.41588	10.38	0.62	2
sDVEDk	751.23096	44.97	0.58	1

The second analysis was performed on October 20, 2017. SEQUEST indicated an ArsR protein score of 0.00 and a percent coverage of 26.22%. The 8 peptides matched from ArsR were of low confidence, and they were also completely different from the first set of peptides identified from the initial sample run. The sample was run again to see if any consistency could be obtained.

Table 3.9 Shotgun Sample Analysis, 10/26

Sequence	MH+ [Da]	RT [min]	XCorr	Charge	Corresponding Sequence in Other Analyses and Analysis Date
ELLARIQSLLR	1310.27854	8.77	1.33	3	eLLARIQSLLRR,10/12
iQSLLRRSHk	1295.82676	0.97	1.09	3	IQSLLRr, 10/12
QKHIPIIISSAR	1363.30973	33.12	0.98	3	QKHIPIIISSAR, 10/20
lRSKIEK	916.59673	43.19	0.83	2	lRSKIEK, 10/20
AEYEILSLLISKk	1521.69219	19.95	0.63	2	AEYEILSLLISK, 10/12
EEVSEPGDANIFr	1479.02056	16.59	0.54	2	KEEVSEPGDANIFR, 10/12
sDVEDk	750.89716	39.34	0.49	1	SDVEDk, 10/20

The third analysis took place on October 26, 2017. SEQUEST scored the ArsR protein with 0.00 and reported a percent coverage of 29.33%. The 7 peptide sequences above were of low confidence, but each of them were identified in the first or second analysis with some slight differences in modification or size of the peptide matched.

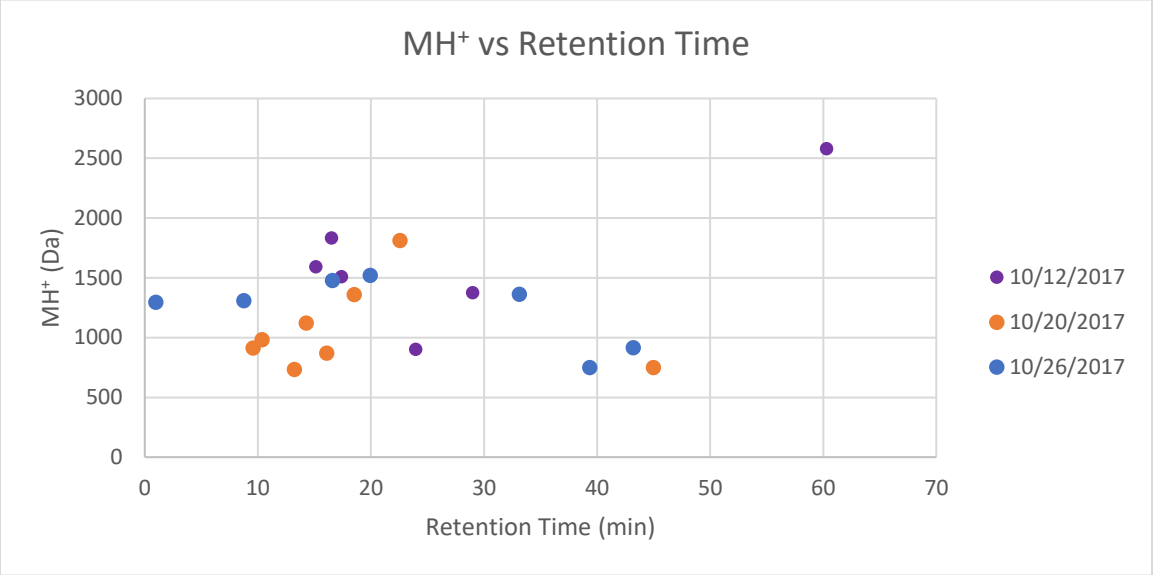

Figure 3.26 Peptide MH^+ vs Retention Time, Shotgun Sample Analyses

Above, each run's identified sequences of ArsR are plotted by their MH^+ values vs their retention times, showing lots of elution prior to 20 minutes, as with the purified protein samples. It can also be seen that, despite the peptide sequence congruence with other analyses to those identified on 10/26, modifications and peptide length affect retention times. Additionally, despite identifying the same sequence and modifications, the experimental fragmentation spectra were not always occurring at the same retention times. For example, lRSKIEK, identified on 10/26, had a retention time of 43.19 minutes, but lRSKIEK, as identified on 10/20, had a retention time of 9.59 minutes. The confidence for both of these matches were low, meaning that SEQUEST could have been incorrect in its identifications. The ion counts for the ions used to match the experimental

spectra to theoretical spectra for each sequence identified as deriving from ArsR were under 100 for all but three ion peaks.

Chapter 4: Nanospray and Electrospray Methods Development

As previously discussed, nanospray can drastically improve ionization efficiency in comparison to electrospray. An NSI source operates at low flow rates of nanoliters per minute, spraying small droplets from a tapered orifice with an inner diameter of as small as 1µm (*28*). The charge-to-volume ratio is much higher originating from NSI than from ESI due to smaller initial droplets (*28*, *29*). Since this is the mechanism for ion formation - coulombic explosions as like charges repel each other and the solvent evaporates – the ionization efficiency increases with NSI (*28, 29*). The desolvation efficiency also increases due to the small, uniformly sized drops (*28*). ESI creates larger droplets that must undergo more fissions before ions are freed (*29*). More solvent must evaporate, and as more droplets are produced, concentration increases, causing problems with salts that compete for ionization (*29*). NSI's smaller initial droplets allow a higher resistance toward salt effects in sample solutions (*29*).

Because of these potential benefits, especially for working with complex shotgun proteomic samples, a PicoView PV-550 nanospray source (New Objective) was installed and tested thoroughly in attempts to optimize it for experimental use.

4.1 Nanospray Emitters

New Objective PicoTip EcontoTip emitters of a 1 ± 0.5µm tip inner diameter (ID) were tested first. These tips were pre-coated in order to make electrical contact for the electrostatic dispersion of sample solution (*28*). To install them into the NSI source, they were mounted in the coated tip module (Fig. 4.1).

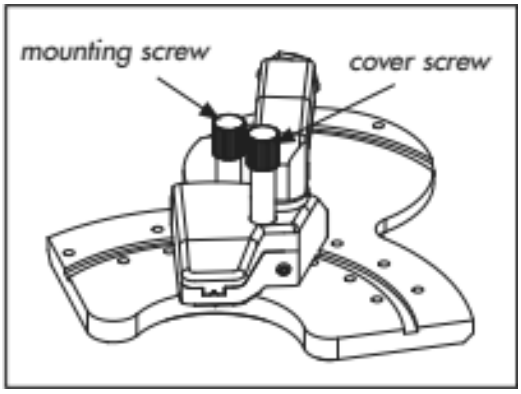

Figure 4.1 Coated Tip Module Mounted in
the Forward Position (*43*)

These tips were made of glass and the manufacturer recommended their off-line use with flow rates of 20-80nL/min, but these were used initially to test the NSI source at about 300 nL/min online with HPLC.

Emitters needed trimming prior to mounting. This was performed with a ceramic wafer. The tip would be scored deeply and snapped at the score. It was not feasible to break these by only tension. This produced jagged edges that made it difficult to get tight seals in unions with fused silica capillary. Additionally, being made of glass and pulled to a very fine, tapered tip, the emitters were incredibly fragile. Throughout the process of installation, these needle tips were frequently broken. The tip mount has an indentation where the union for the emitter and the emitter tip fits.


Figure 4.2 Union of Fused Silica Transfer Line
and Emitter in Coated Tip Module (*43*)

Putting the needle into the mount and removing it was risk for breakage. While testing seals and re-mounting, the coating was relatively easy to remove. A 0.5 mL gas-tight syringe was used to push solvent through the capillary, checking for leaks and for visible spray, but contact with Solvent A (98% Ultrapure H_2O, 2% ACN, 0.2% formic Acid) was especially damaging to the coating. As this was realized, gloves were worn, tweezers were used for handling, and care was taken to limit the metallized coating's exposure to liquids. Much time was spent attempting to install these emitters with little continuous spray result.

Often, when no leaks were observed, and needles were properly installed, there would still be no visible spray. This indicated clogging, which could sometimes even be seen via the CCD camera. It was gathered that scoring with a wafer cutter and snapping emitters was leaving glass debris that found its way into the fine tapered tip. Additionally, any standard solution being tested contributed to needle clogging as well. Once a tip became clogged, there were attempts made to unclog using sonication, but usually this was the end of the lifetime of an emitter. Sonication was a delicate process, just like tip installation. The tips had to be submerged in liquid; both Solvents A (composition above) and B (composition: 98% ACN, 2% H_2O, 0.2% formic acid) were tried. This submersion usually removed some of the tip coating, especially for the glass EconoTips. Additionally, it was realized that the tip could not be placed in the bottom of a beaker, for the fine point would easily be lost. Parafilm was used to attach and hang the emitter from the beaker opening, only submerging about 1/3 of the emitter in solvent. While this loosened debris sometimes, and spray could be achieved for a few moments after re-installation, the particles would soon clog the tip again. When spray was accomplished for long enough to attempt a sample run, emitters were easily clogged with sample contents and their contaminants as well. The methods for HPLC and MS/MS for NSI runs are detailed in sections 4.3 and 4.4.

In addition to proper installation into the coated tip module, the module had to be fixed to the magnetic stage plate with a mounting screw (Fig. 4.2). Then the stage plate was attached to the translation stage, ensuring electric grounding by attaching a metal union to a grounding clip on the source, then the source would be attached to the MS interface initiating the flow of electricity for spray voltage. Next, the tip would be adjusted in the X, Y and, Z directions to position it at the MS inlet. Again, throughout this stage plate attachment and adjustment process, the emitters would break, hitting the inlet, another part of the source, or a person.

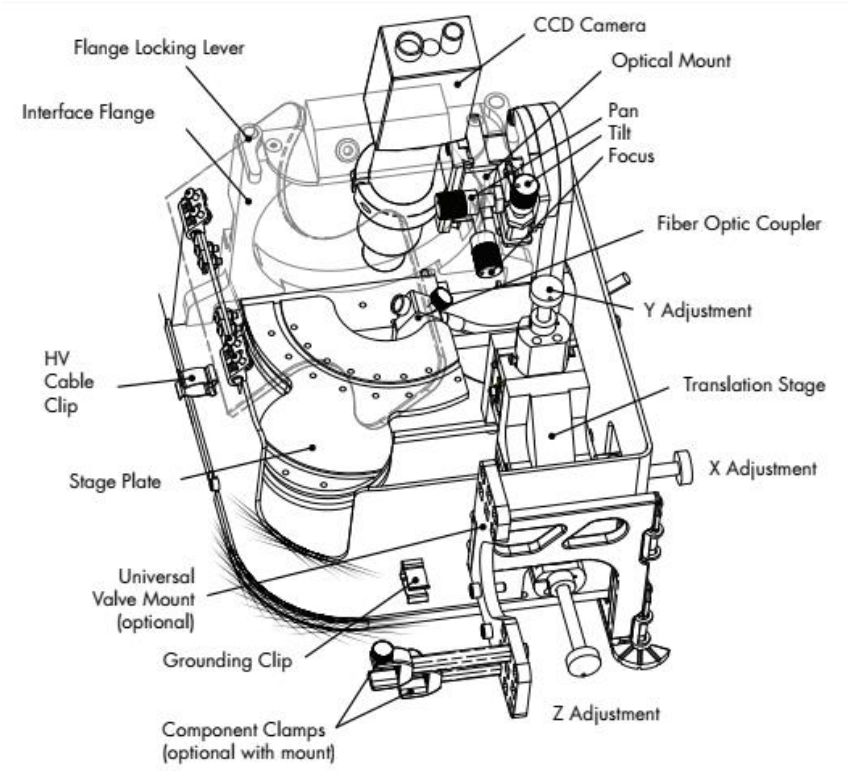

Figure 4.3 PicoView NSI Source User-View (*43*)

At the beginning of summer research, it was realized that the inner diameter of the EconoTip glass emitters was too small to properly accommodate the flow rates for sample analysis; the HPLC method involved a flow split, and during the majority of the gradient time, the flow rate

was 300nL/min (section 4.3). PicoTip SilicaTip emitters were ordered from New Objective. These were coated nanospray needles made from fused silica and had a tip ID of 15 ± 1µm, recommended for flow rates of 200-500nL/min. Before these emitters came in, another New Objective SilicaTip was available for testing, but only one. This emitter had a tip ID of 30 µm and was recommended for flow rates between 300 and 1000nL/min. This tip was used for runs of a shotgun E. coli/T7 Phage sample (another student's proteomics project) and two Solvent A blanks. Clogging was observed after the first blank run. The tip was sonicated, and spray was observed, so another blank was run, but the signal was lost during (section 4.5)

Soon after, the new emitters arrived. These coated emitters were much easier to work with than the glass ones; their flexibility allowed them to be shortened by scoring and pulling, rather than snapping, and their tips were less prone to accidental breakage. The coating on these needles was also more resilient to solvent contact. Tight unions were accomplished more easily, and emitters sprayed upon installation. To attempt to extend the lifetimes of the first two of these needles, the HPLC and MS were not turned off except to load a sample into the HPLC tray. When samples or blanks were not being analyzed, a low flow rate (20 µl/min) of solvent A was constantly supplied. This seemed to only have slight effect on emitter lifetime. Purified ArsR samples were run. One of these runs produced promising results, with high protein scoring in SEQUEST, but indicated the need for analytical column replacement (Section 4.5)

At this point, it was confirmed that NSI emitters are typically one-time use with lifetimes lasting up to about 2 days at most, but usually only a few hours (*44*). Their tips are prone to clogging, and their conductive coatings are susceptible to degradation (*45, 46*). Additionally, they would be used only once to prevent sample contamination; a blank run is generally not feasible for such frequent clogging (*47*). There is a high cost for commercially produced NSI emitters, but they

can potentially clog before a successful LC/MS run. Because only three (of a total of five) of the new SilicaTip emitters remained, literature was consulted to learn how to pull NSI tapered emitters in-house, but many researchers utilized laser-heated pulling (*45*, *48*) or other techniques that were not viable (plasma-pulling (*47*), HF etching (*46*, *48*)). It was decided that a simple, inexpensive, heating and pulling approach would be taken (*49-51*).

A butane torch, fused silica tubing, tape, and a binder clip were utilized to make emitters.

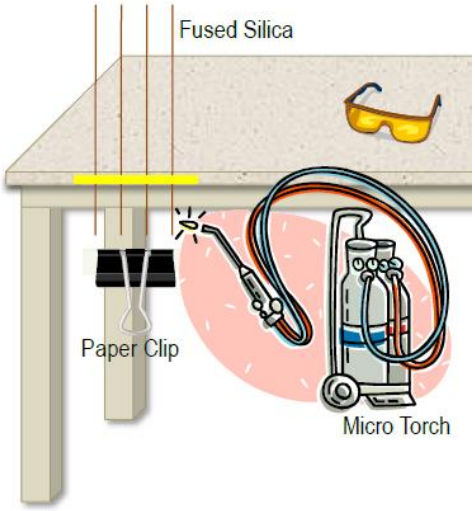

Figure 4.4 NSI Tip Pulling (*51*)

To practice, the available tubing was used, which included both 75 µm ID x 358 µm OD and 254 µm ID x 360 µm OD fused silica tubing coated in polyimide. The tubing was cut with a silica wafer cutter into pieces of about 10 cm in length and taped to the edge of a desk. A binder clip was gently attached to the ends of the three or four pieces were taped down at a time. Then, a butane torch was used to heat them until the coating began to burn and the silica melted. The binder clip acted as a weight, pulling the silica to fine, tapered points as it melted. The pulled tips were examined under a stereoscope and compared to the 15 µm ID tips of the New Objective SilicaTips to determine where cuts should be made. Attempts to use the wafer cutter to score the tips were mostly unsuccessful. The tips were easily crushed or broken because they were thin and fragile. It

was difficult to cut or break the ends at the point where the taper had the correct ID. The jagged edges of breaks could be seen underneath the stereoscope. Unclean edges at the tapered tips of emitters are a hindrance to proper nanospray. The other ends were examined and cut as well so that the needles were about the same length as the shortened Silicatips that had already been fit into the NSI source mount. These ends were jagged as well preventing tight seals in unions. A few days were spent trying to perfect the cutting technique. It was clear that the wafer cutter would not suffice for clean cuts, so other cutting options were researched. Diamond scribes were ordered and used after this. A shortix cutter was also ordered, but was not able to be used before the source was switched back to ESI.

To test pulled emitters for spray, the uncoated tip module was used. There are two possible set ups for this module, a forward and a back position. The forward position is recommended for uncoated, unpacked, tapered tips, but the voltage is supplied by a microtee with a platinum electrode, and the PicoView manual states that the microtee is unsuitable for low-volume, post-column use. Its interior volume introduces dead volume after the column in which peaks broaden. A zero dead volume titanium union can be purchased, but was not for these initial tests. The forward position was tried briefly, but when no spray was observed, the back position was used. The back position is recommended for PicoFrit columns (nano HPLC column packing within the emitter). If a sample had been run, the EASY-column (section 4.2 and 4.3) would have been placed in the microtee and supplied with high voltage through the liquid junction and the pulled tips would have been mounted in front.

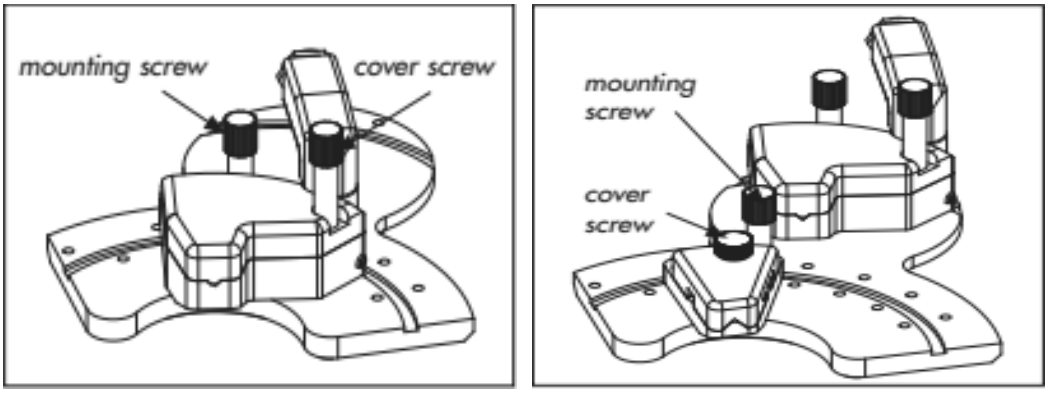

Figure 4.5 Uncoated Tip Module Mounted in the Forward Position (Left) or in the Back Position with the Tip Holder Forward (Right) (*43*)

Since the needles were just being tested for spray by way of a syringe pump, a fused silica transfer line was connected to one side of the electrode and another fused silica capillary was connected to the other side where the EASY-column would have been for a sample run. The intermediate capillary was joined to the emitter in the tip holder by a MicroTight (UpChurch Scientific, IDEX Corporation) union. Since the uncoated tip module had never been used before, a set of uncoated New Objective Silicatips (15µm ID, 75µn OD) were ordered for testing as well, but these did not arrive until late in the summer when ESI was reinstalled.

Figure 4.6 Microtee with Platinum Electrode Loaded in to Uncoated Tip Module in the Back Position (*43*)

A variety of tips were pulled and tested. To check for spray, solvent A was used. If spray was observed, a 10^{-5} M bradykinin solution (20% ACN, 0.1% Formic Acid) was employed to check signal. Since most of these needles were pulled from fused silica of 254 ID and 360 OD, the tips ended up with a larger inner diameter than the Silicatips, and higher flow rates were used to test their spray (between 450nL/min and 700nL/min). Spray voltages were between 1.5kV and 2.0kV. As more needles were pulled and tested, spray was observed more often, and for one tip, bradykinin peaks were observed with signal intensity at 10^6 or 10^7 (section 4.5). Spray typically lasted no longer than a few minutes. 20 minutes was the maximum observed. A 180 m/z peak was consistently present, especially when the bradykinin peaks were not being observed from solution. It was hypothesized that this might be a result of the residual coating. For the last batch of emitters made, the polyimide coatings were burned and wiped off with methanol prior to pulling. The emitters tested from this batch were less successful than the other tips, and ESI was replaced before the entire batch was tested. It should be noted that the coating gives the capillaries their strength and flexibility. This group of emitters was very easy to break during installation.

4.2 Column Packing

Just like NSI emitters, nano HPLC columns are expensive to purchase. The two nano columns used during NSI experimentation were purchased from Thermo Fisher Scientific. They were 10 cm long EASY-columns with a 75 µm ID and C18 packing material of 3 µm. After replacing the first column, a column packing station was set-up and tested.

First, fritted capillaries were made. 75 µm ID by 358 µm OD fused silica capillary was cut into a total of five pieces, each 25 cm in length. Next, 200µl of KASIL 1 potassium silicate solution (29.1% solution in water) was added along with 50 µl of formamide (Fisher Scientific) to a 1.5 ml Eppendorf tube. This solution wasn't vortexed before centrifuging, but the next time this procedure

is performed, it should be mixed more thoroughly. The tube was centrifuged at 10,000 rpm for 2 minutes. Next, the ends of four of the capillaries were dipped, one at a time, into the solution for 5-10 seconds each. The solution traveled up the tubes to about 2 cm and were examined by stereoscope to check for the length of the frit and for air bubbles. 50 µL of solution was also dripped onto a piece of filter paper. The fifth capillary's end was touched to the filter paper spot. Upon inspection, this method did not seem to work, and the fifth capillary was also dipped into the solution as the other four had been. The capillaries were placed into an oven at 90˚C overnight. They were removed from the oven the next morning, allowed to cool, and inspected under a stereoscope. Three of the five looked to have successful frits.

Next, a column packing station was set up. This required a helium tank and a pressure cell with a three-way valve. A high pressure regulator was attached to the helium tank because packing occurs at 1000 psi. Stainless steel tubing connected the helium to a three-way valve that was connected to the pressure cell. Connections were made using Swagelok fittings. Because an Eppendorf tube filled with ACN or packing material needed to sit upright at the bottom of the pressure cell chamber, a brass holder was machined in the physics shop to fit snugly in the cavity of the chamber and hold the tube. It did not have enough height to reach the midsection of the Eppendorf tube, so a washer was placed underneath it.

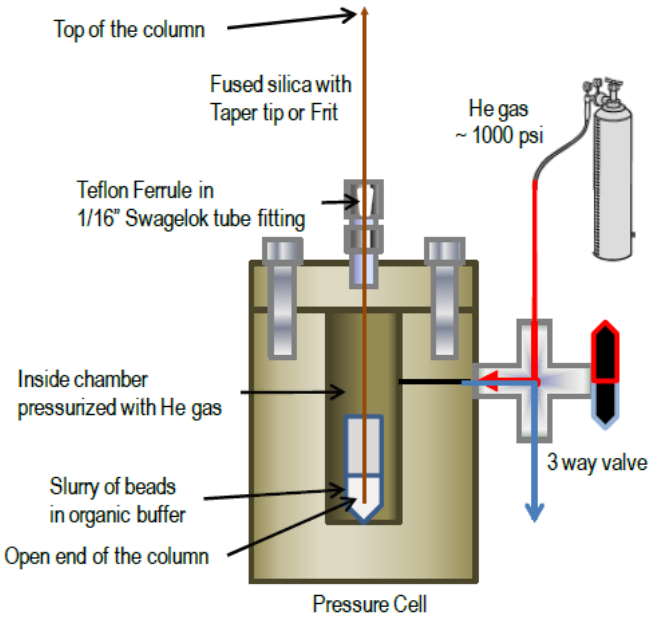

Figure 4.7 Brechbeuhler-Type Pressure Cell for Column Packing (*51*)

To equilibrate the capillary and test the pressure cell arrangement, an Eppendorf tube was filled about two-thirds of the way up with acetonitrile, the lid was cut off, and it was placed in the bottom of the cell. The lid of the cell was secured by washers and bolts, and the capillary (open/non-fritted end) was threaded through a Swagelok tube fitting, a PTFE ferrule (1/16’’), and through the lid. The open end was positioned so that it would be in the ACN but not touching the bottom of the tube. The nut was secured by hand. Tightening with a wrench would have crushed the ferrule. Next, the pressure on the regulator was set to less than 1000 psi, for safety, and the gas-flow was turned on. The three-way valve was opened, and the pressure was slowly increased to 1000 psi. A drop of ACN was noticed at the fritted end of the capillary, and it was permitted to equilibrate for about 5 minutes. After 5 minutes, the gas was shut off, and the three-way valve was closed. Before removing the lid and the capillary, the cell was vented using the three-way valve until no more gas could be heard escaping. This lid was removed, and the capillary was the ready

to be packed. Column packing did not take place because shortly after this, the ESI source was re-installed and nano HPLC was no longer in use.

4.3 High Performance Liquid Chromatography (HPLC) Methods and Plumbing

Originally, the HPLC was plumbed to include a precolumn (Acclaim PepMap 100) that helped prevent contaminants from reaching the analytical column and the NSI source. The precolumn had an ID of 300 µm, a length of 5mm, and C18 particle sized 5 µm. The gradient for the 100-minute precolumn method can be viewed below. It follows the same changes in percentage of B over time as the 100-minute direct method which was described in section 2.2.

Table 4.1 HPLC Gradient 100-Minute Precolumn Method, NSI

	Time (min)	Event	Parameter
1	0.01	Start	
2	0.02	Total Flow	30 µL/min
3	0.10	Pump B Concentration	2%
4	5.00	Total Flow	30 µL/min
5	5.01	Total Flow	300 µL/min
6	5.02	Pump B Concentration	2%
7	60.00	Pump B Concentration	55%
8	60.10	Pump B Concentration	90%
9	70.00	Pump B Concentration	90%
10	70.10	Pump B Concentration	2%
11	99.00	Total Flow	300 µL/min
12	99.10	Total Flow	30 µL/min
13	100.00	Stop	

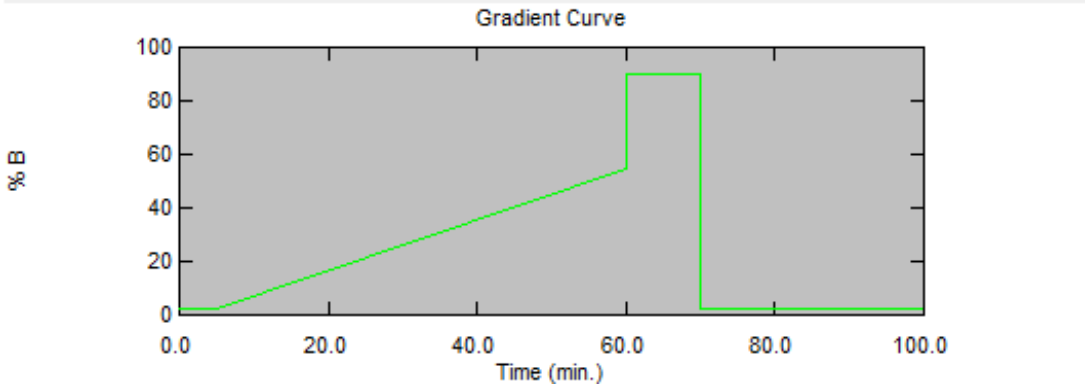

Figure 4.8 HPLC Gradient Curve 100-minute Precolumn Method, NSI

In the Shimadzu UPLC, a flow-splitter was present to allow the use of nanoflow rates. A split of 1000:1 was used. The flow was split after gradient mixing and before the autosampler. During the first 5 minutes of a run, 98% A and 2% B flowed from the gradient mixer, through the autosampler, and into the trap column at about 30 µL/min. The sample (if sufficiently hydrophobic) was attracted to the C18 packing material, while the rest of the flow went to waste. At 4.85 min, the divert valve switched, and the mobile phase then flowed back through the trap column. This is where the low-flow line began to be utilized for sample movement. The flow rate increased overall at 5.01 minutes to 300 µL/min, meaning that the low flow rate was about 300 nL/min. From there, the gradient progressed as usual, and more non-polar analytes were washed off the column in time as the percentage of organic solvent increased and analytes' affinities for the mobile phase increased. A depiction of the divert valve and the system before and after the switch can be viewed below.

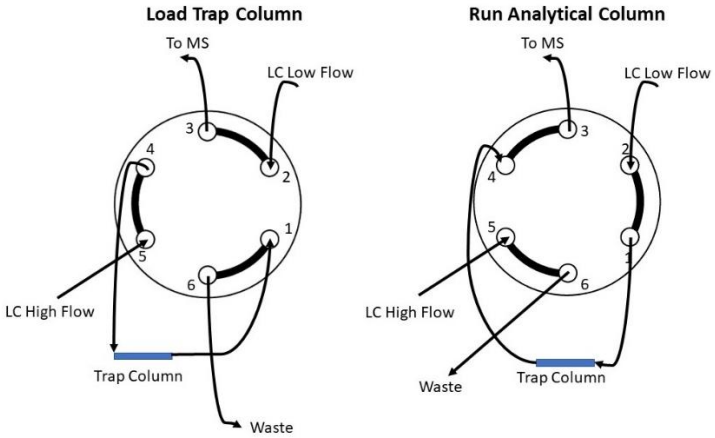

Figure 4.9 Divert Valve

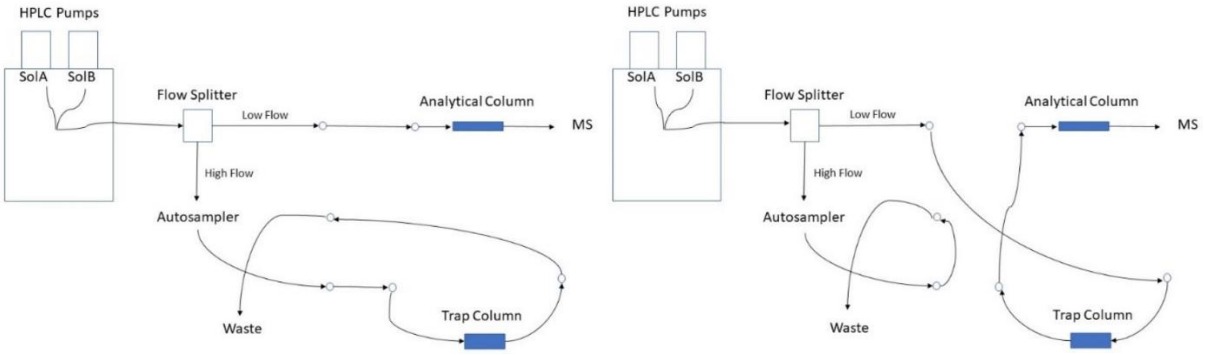

Figure 4.10 Loading Trap Column (Left) and Running Analytical Column (Right)

When the first EASY-column was replaced for another, the gradient was extended to 170 minutes because it was noticed that there was an abundance of peptides identified by SEQUEST that eluted at the end of the gradient when the percentage of solvent B was at 90%. The goal of the lengthened gradient was to increase the amount of time at high percent solvent B to try to get more peptides off the column.

Table 4.2 HPLC Gradient 180-Minute Precolumn Method, NSI

	Time (min)	Event	Parameter
1	0.01	Start	
2	0.02	Total Flow	30 µL/min
3	0.10	Pump B Concentration	2%
4	5.00	Total Flow	30 µL/min
5	5.01	Total Flow	300 µL/min
6	5.02	Pump B Concentration	2%
7	120.00	Pump B Concentration	55%
8	120.10	Pump B Concentration	90%
9	140.00	Pump B Concentration	90%
10	140.10	Pump B Concentration	2%
11	169.00	Total Flow	300 µL/min
12	169.10	Total Flow	30 µL/min
13	170.00	Stop	

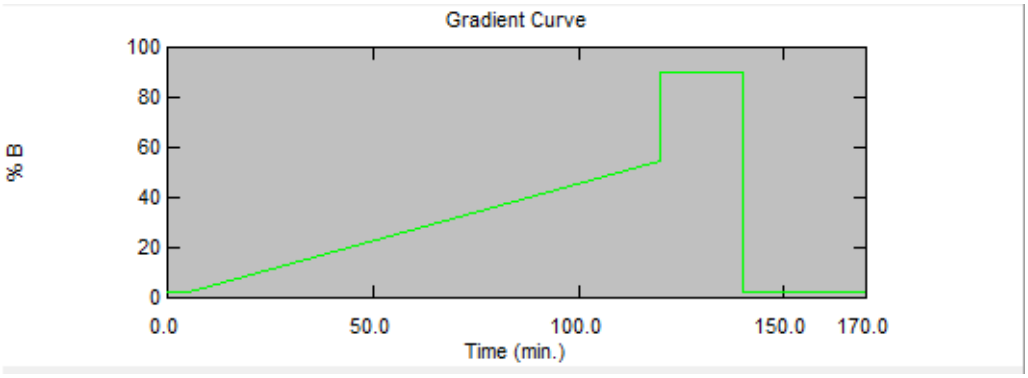

Figure 4.11 HPLC Gradient Curve 170-Minute Precolumn Method, NSI

Without a change in results, the original 100-minute method was adopted again.

It was clear that the first purified ArsR sample was prepared correctly based on SEQUEST's scoring from a run utilizing a New Objective coated SilicaTip (section 4.5), but it was unclear why results were inconsistent. When run after run was not providing optimal SEQUEST results, the system was reconsidered. The trap column was removed to attempt to reduce sample loss, increase signal, and increase SEQUEST scoring and identifications. It was thought that maybe the sample wasn't binding sufficiently to the precolumn, so its bypass would help analytes make it to the NSI source and into the MS.

The LC system was replumbed to avoid the trap column, and the divert valve was turned off in the MS method. The high flow line was taken directly to waste, and the low flow line went to the autosampler and then to the analytical column and the MS. For the entirety of the new 100-minute direct gradient, the total flow rate was 300 µL/min, meaning that the low-flow line in which the sample was injected, operated at 300 nL/min. The same changes in percentage of solvent B over time were utilized for this gradient.

The gradient wasn't altered again until the flow splitter was removed, and NSI was switched out for ESI. The 100-minute ESI direct HPLC method was the same as that for NSI except the total flow throughout the gradient was 400 µL/min. A shorter gradient of 60 minutes

was tried as well, and another 60-minute gradient was altered to extend the region in which the percent B concentration was high. A 400 µL flow rate was held constant across methods. Instead of rising to 55% then jumping to 90%, the percent B rose to 75% and then went to 90% remaining there for 10 minutes. The original 60-minute direct method only remained at 90% solvent B for 5 minutes. The 60-minute method and the altered 60-minute method gradient curved can be viewed below.

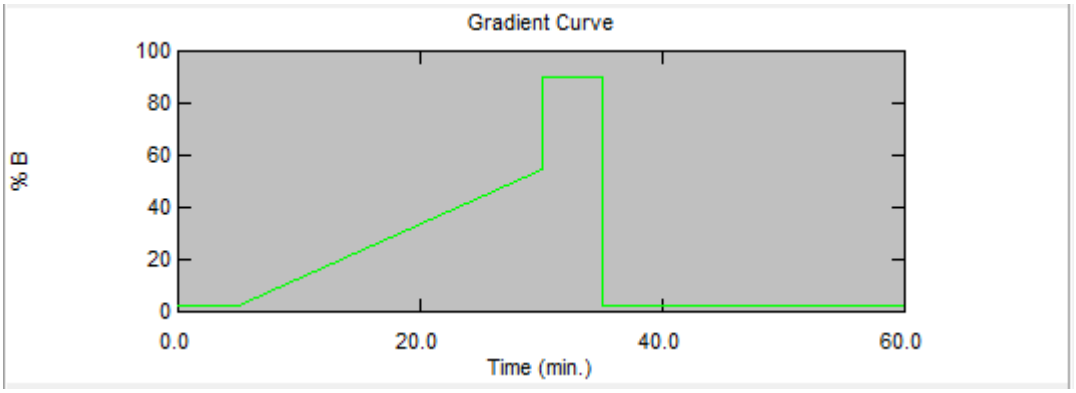

Figure 4.12 HPLC Gradient Curve 60-Minute Direct Method, ESI

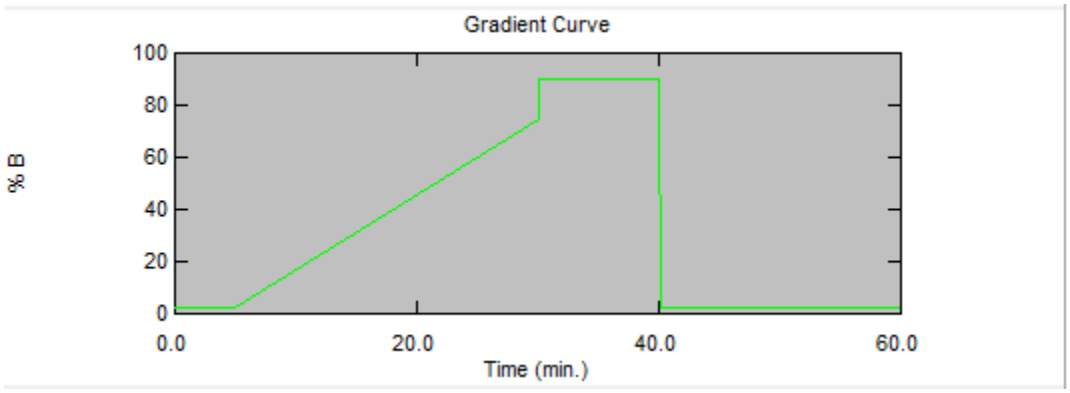

Figure 4.13 HPLC Gradient Curve Altered 60-Minute Direct Method, ESI

While trying to optimize the ESI after installation, standard solutions, an LC-MS calibration solution, a peptide mix, and the 10^{-5} M bradykinin solution, and samples were run with 100-minute and 60-minute gradients (section 4.5).

4.4 Data-Dependent Mass Spectrometry Methods

Besides spray voltage, sheath gas, and, specifying divert valve positions, the 100-minute MS/MS methods coupled to the NSI source were the same as those described for the ESI source (section 2.3). The sheath gas flow rate was set to 8 arbitrary units. For the precolumn method, the divert valve position was to the source at 0.00 minutes and switched to the waste position at 4.85 minutes. The source position aligns to the "load trap column" position in Figure 4.9. The waste position is analogous to the "run analytical column" position of the same figure. The spray voltage varied. At the beginning of the summer, the spray voltage was high (2.8 kV) for a nanospray source. This reflected the fruitless attempts to spray from the glass EconoTip emitters. Gradually, it was discovered that NSI spray voltages are much lower than ESI for stable spraying (*43*). Voltages above 4 kV is excessive, but 4 kV was applied to the EconoTips, and corona discharges were witnessed even in the 3 kV ranges. A factor in the inability to see sample in spectra, high spray voltages may have led to the desorption of coating or contaminants off the tip of emitters. This idea is supported by the tests of tips in which spray was observed visually, but standard peaks were not observed in the spectra. NSI may only require 500 volts for proper spray and ionization (*44*). The PicoView source manual makes mention of 25 µm PicoTips requiring 2 kV or more (*43*). In literature, it was seen that a range of voltages from about 450 to 2500 was successfully applied to NSI tips (*28, 44, 46-48, 50, 52, 53*). Once this was grasped, spray voltages were applied within a range of 1.5 kV to 2.5 kV for emitter testing and sample analysis. For ESI, the spray voltage was held at 5.0 kV.

4.5 Results and Discussion

A total of three samples were analyzed using NSI while the emitters were functioning correctly: an *E. coli* sample on 05/23/17 with the 30 µm ID SilicaTip, a purified ArsR sample on 06/06/17 with a 15 µm ID SilicaTip, and a second run of the same purified ArsR sample on 06/07 with the same 15µm ID SilicaTip. These three runs, along with the blanks that followed them, scored highest in SEQUEST when compared to any other proteomic sample analyses performed, NSI or ESI. All three samples were analyzed with the 100-minute precolumn method.

Proteins | Peptides | Search Input | Result Filters | Peptide Confidence | Search Summary

	Accession	Description	Score	Coverage	# Proteins	# UniquePeptides	# Peptides	# PSMs	# AAs	MW [kDa]	calc. pI
1	A0A140N4H5	Autonomous glycyl radical cofactor OS=Escherichia coli (s...	2157.29	66.93 %	1	6	10	1245	127	14.3	5.17

	A2	Sequence	# PSMs	# Proteins	# Protein Groups	Protein Group Accessions	Modifications	ΔCn	XCorr	Charge	MH+ [Da]	ΔM [ppm]	RT [min]
1		aGYAEDEVVAVSK	2	1	1	A0A140N4H5	N-Term(Acetyl)	0.0000	3.28	2	1378.36748	-944.22	57.14
2		AGYAEDEVVAVSK	131	1	1	A0A140N4H5		0.0000	3.19	2	1338.27239	458.79	61.45
3		FNSLTPEQQR	132	1	1	A0A140N4H5		0.0000	3.19	2	1219.60173	-4.01	75.54
4		MITGIQITK	119	1	1	A0A140N4H5		0.0000	2.84	2	1005.10210	518.52	58.80
5		EVPVEVKPEVR	145	1	1	A0A140N4H5		0.0000	2.69	2	1281.38286	516.56	58.69
6		LGDIEYR	127	1	1	A0A140N4H5		0.0000	2.66	2	865.93730	572.63	62.17
7		mITGIQITK	101	1	1	A0A140N4H5	M1(Oxidation)	0.0000	2.59	2	1021.04839	462.80	82.58
8		VEGGQHLNVNVLR	128	1	1	A0A140N4H5		0.0000	2.48	3	1437.26493	1728.06	56.58
9		YPQLTIR	129	2	2	A0A140N4H5;A0A140NBE7		0.0000	1.97	2	890.77178	294.49	70.23
10		RETLEDAVK	76	1	1	A0A140N4H5		0.0000	1.94	2	1060.69695	125.93	76.25
11		ETLEDAVK	105	1	1	A0A140N4H5		0.0000	1.65	2	905.03557	633.48	66.71
12		TFTESL	48	1	1	A0A140N4H5		0.0000	1.33	1	697.25006	-129.45	91.82
13		IGDIEYR	2	1	1	A0A140N4H5	N-Term(Acetyl)	0.0000	1.11	2	906.34673	-1219.49	65.66

	Accession	Description	Score	Coverage	# Proteins	# UniquePeptides	# Peptides	# PSMs	# AAs	MW [kDa]	calc. pI
2	A0A140NEN6	10 kDa chaperonin OS=Escherichia coli (strain B / BL21-D...	784.17	47.42 %	1	3	4	326	97	10.4	5.24
3	A0A140NCH9	DNA-binding protein OS=Escherichia coli (strain B / BL21-...	700.65	43.07 %	1	2	8	295	137	15.5	5.47
4	A0A140NF24	Histone family protein DNA-binding protein OS=Escherichi...	627.09	33.33 %	1	2	3	203	90	9.5	9.58

Figure 4.14 *E. coli* Shotgun Sample, SEQUEST Report Example

The goal of the *E. coli* project was to look at the differential expression of proteins over time after T7 phage infection. The highest scoring protein in the SEQUEST report for the *E. coli* sample run scored 2157.29. Lower on the list, proteins still have scores in the hundreds. Many of the peptide sequences matched from the autonomous gylcyl radical cofactor protein had hundreds of spectral matches. The highest xCorr was for this protein was 3.28, and all but two peptide sequence matches had Xcorr values of over 1.50. The experimental product ion spectrum matched with the highest Xcorr to aGYAEDEVVAVSK had ion counts in the thousands.

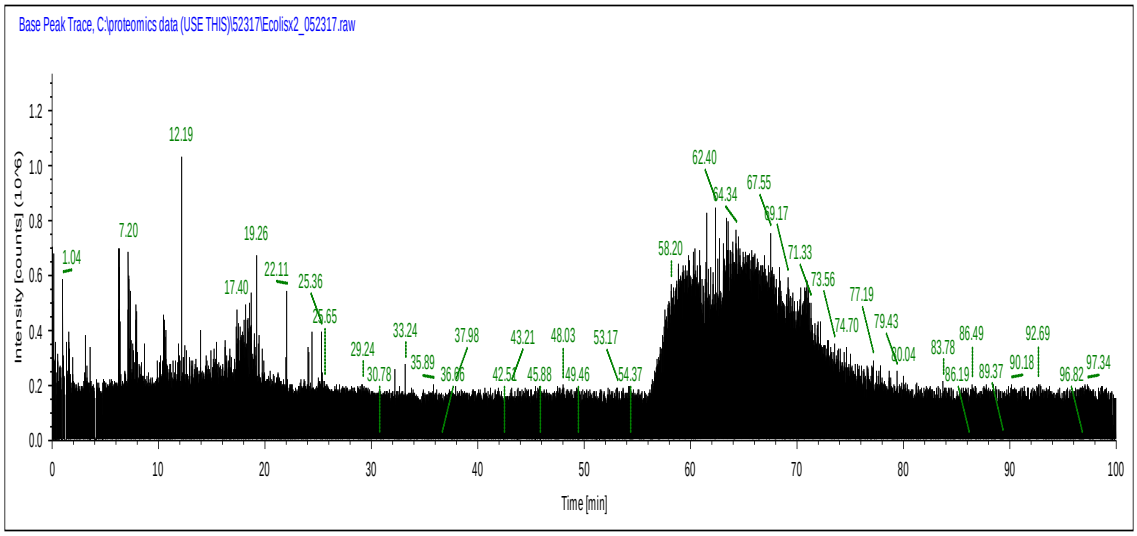

Figure 4.15 Chromatogram of Shotgun *E. coli* sample, 100-Minute Precolumn Method, NSI, 2.8 kV

To spray sample solution, a 2.8 kV potential was applied throughout the LC-MS run. The emitter was functioning correctly, so the chromatogram could be analyzed with more depth. The peaks do not appear to be clearly resolved, and large amounts of sample and contaminants washed off the column after about 55 minutes of gradient. This indicates that the analytical column needed replacing.

A solvent A blank was run after the *E. coli* sample. When analyzed against the E. coli and T7 FASTA file, the highest scores for proteins remained in the hundreds. The highest scoring protein was again the autonomous glycycl radical cofactor with a score of 593.85.

oteins | Peptides | Search Input | Result Filters | Peptide Confidence | Search Summary

	Accession	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# AAs	MW [kDa]	calc. pI
1	15644795	response regulator ompR [Helicobacter pylori 26695]	5243.63	64.44 %	1	13	24	2830	225	25.8	5.43

	A2	Sequence	# PSMs	# Proteins	# Protein Groups	Protein Group Accessions	Modifications	ΔCn	XCorr	Charge	MH+ [Da]	ΔM [ppm]	RT [min]
1		HIPIIISSAR	396	1	1	15644795		0.0000	2.28	1	1106.85681	170.49	59.94
2		ALDYGADDYLPKPYDPK	318	1	1	15644795		0.0000	3.36	2	1941.61528	354.13	82.59
3		KEEVSEPGDANIFR	308	1	1	15644795		0.0000	5.03	3	1592.56638	1124.27	58.64
4		SHKKEEVSEPGDANIFR	143	1	1	15644795		0.0000	3.70	2	1944.40483	742.14	78.28
5		GIGYKLEY	238	1	1	15644795		0.0000	2.77	2	943.26982	823.40	97.55
6		ESIAIESESINPESSNK	143	1	1	15644795		0.0000	5.30	2	1835.81377	1058.08	52.41

Figure 4.16 Purified ArsR Sample 1, 06/07, SEQUEST Report Example, NSI

The use of NSI to ionize and vaporize purified ArsR resulted in a SEQUEST score of 5243.63 and a sequence coverage of 64.44% for response regulator ompR (ArsR). 28 peptides were matched deriving from the ArsR sequence: 11 high confidence, 2 medium confidence, 15 low confidence. All of the high confidence sequences had hundreds of experimental spectra matches. The highest Xcorr value for a peptide sequence match was 5.30. 16 sequences had Xcorr values above 1.50.

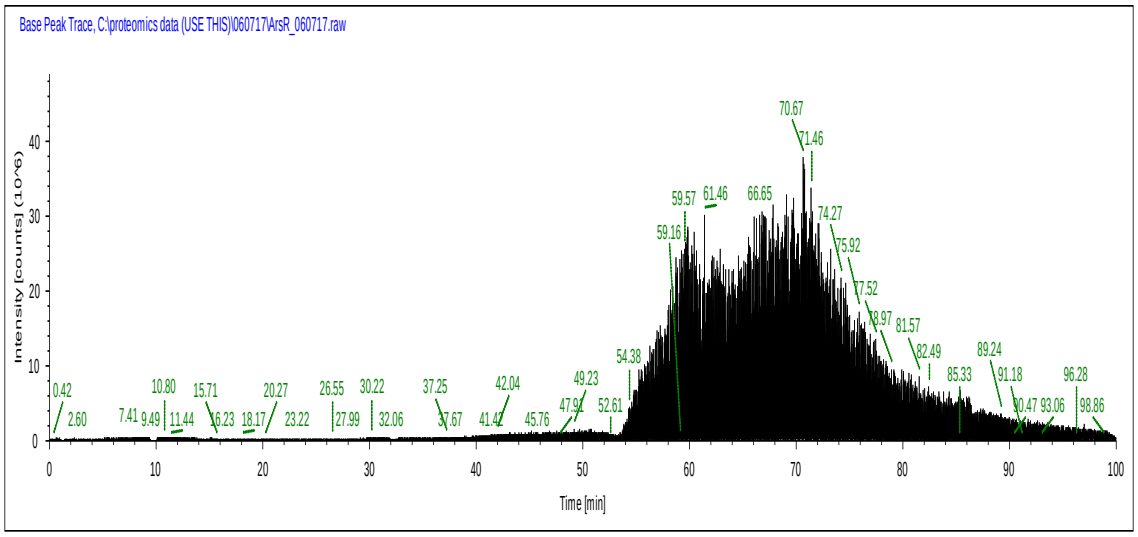

Figure 4.17 Chromatogram of Purified ArsR Sample 1, 100-Minute Precolumn Method, NSI, 2.1 kV, 06/07

Just as with the *E. coli* shotgun sample, the functioning of the emitter allowed for true evaluation of other aspects of the LC-MS setup. The large amount of elution after 55 minutes indicates the failure of the column. The solvent A blank run through the LC-MS after this sample also scored highly in SEQUEST for ArsR with a protein score of 1806.80 and a sequence coverage of 54.22%. The results from the 6th of June were comparable to those achieved on the 7th with slightly lower confidence and sequence coverage. Both analyses were performed with the spray

voltage adjusted to 2.1 kV. The protein score for ArsR was 2494.63 with a sequence coverage of 48.44%. 7 of 9 high confidence sequence matches had between 100 and 200 spectral matches.

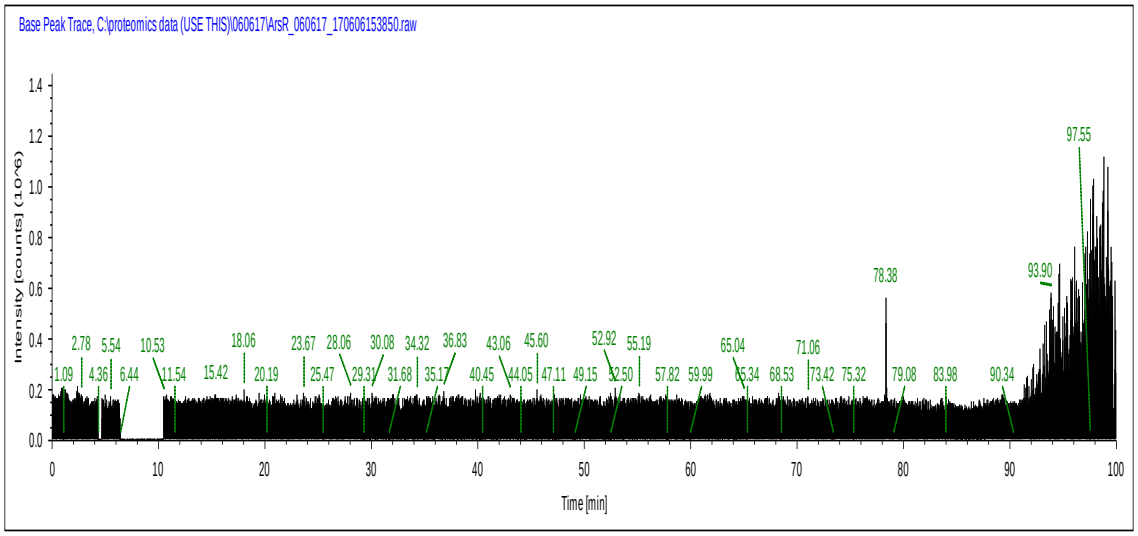

Figure 4.18 Chromatogram of Purified ArsR Sample 1, 100-Minute Precolumn Method, NSI, 2.1 kV, 06/06

Based on the chromatogram, the emitter was not spraying continuously for the first purified ArsR sample run. This likely contributed to the lower scoring in SEQUEST; however, the results are still significantly more confident than all runs that utilized ESI as the ionization source.

The SilicaTip needle was replaced on the 12th of June, but no significant results came of the run of purified ArsR sample 2 with spray voltage at 2.1 kV. The analytical column was then replaced, and a run was performed with the same emitter at a spray voltage of 2.2 kV with a new 170-minute gradient. Again, response regulator ompR (ArsR) was not identified. The spray voltage was increased to 2.5 kV, but still, ArsR was not reported by SEQUEST.

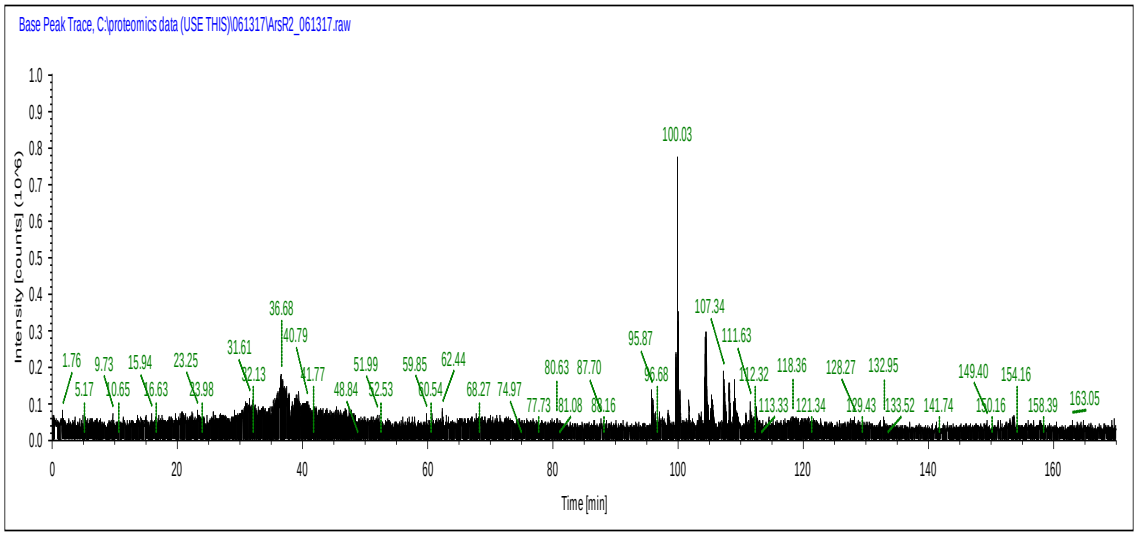

Figure 4.19 Chromatogram of Purified ArsR Sample 1, 170-Minute Precolumn Method, NSI, 2.5 kV, 06/13

There are a few distinct peaks here, but overall, the chromatogram looks very noisy. As mentioned above, ArsR was not identified from this data set.

After another fruitless run with spray voltage at 2.5 kV with the 100-minute gradient, the needle was replaced again. Still, when an *E. coli* sample was analyzed with the 100-minute gradient and a spray voltage of 2.1 kV, nothing significant resulted. Next, the trap column was removed, and a 100-minute direct method began to be applied. A larger sample volume, 50 µL, was injected during the next analysis of purified ArsR. The trap column removal, larger volume sample injection, and 2.4 kV spray voltage yielded no telling SEQUEST results. On 6/20 the needle was replaced again, and a bradykinin solution was used to try to optimize the NSI source. Lower spray voltages were used after this because high voltages seemed to correlate with signal loss.

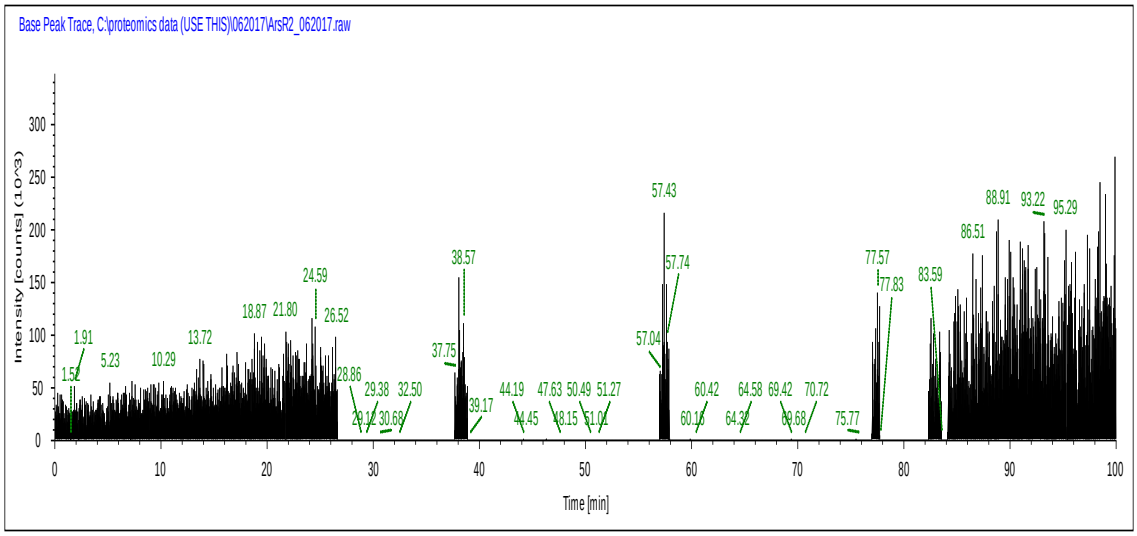

Figure 4.20 Chromatogram of Purified ArsR Sample 1, 100-Minute Direct Gradient, NSI, 1.75 kV, 06/20

As can be seen in the above chromatogram, the emitter did not spray continuously. It may have been damaged by corona discharge or clogged with bradykinin-solution contaminants during attempted optimization. ArsR was not identified by SEQUEST.

With only one New Objective SilicaTip remaining, emitters began to be pulled and tested with 10^{-5} M bradykinin solution.

Table 4.3 Pulled NSI Emitter Tests

	Flow Rate (nL/min)	Spray Voltage (kV)	Spray?	Spray Duration	Bradykinin m/z Peaks Observed (Da)	Bradykinin Max Signal
1	700	1.8	Yes	<1 min	N/A	N/A
2	450	1.5-1.7	Yes	20 min	354, 531, 1060	10^6-10^7 for 354 and 531
3	500	1.5	No	N/A	N/A	N/A
4	500	1.25-2.0	Yes	5 min	531	10^5-10^6 for 531
5	500	1.5-1.8	Yes	*	N/A	N/A
6	500	1.5-1.7	Yes	*	N/A	N/A

* Spray was observed until tip was removed.

Emitters besides the ones listed above were tested as well. First, they were pulled as practice and did not function. Before this batch of needles, it was uncertain that the uncoated tip mount was functioning correctly. Notably, spray was achieved with the above in-house pulled tips. Bradykinin peaks were observed for two of the six listed. The flow rates shown in the table represent those that produced the most stable spray, or in the case of emitter 2, the only flow rate that was attempted. A couple of needles were permitted to spray for an extended period of time, but bradykinin peaks were never observed. Liquid spray was observed emanating from both tips 5 and 6, but bradykinin ions were not entering the MS. Upon their removal, residue was seen in both tips 5 and 6 via a stereoscope.

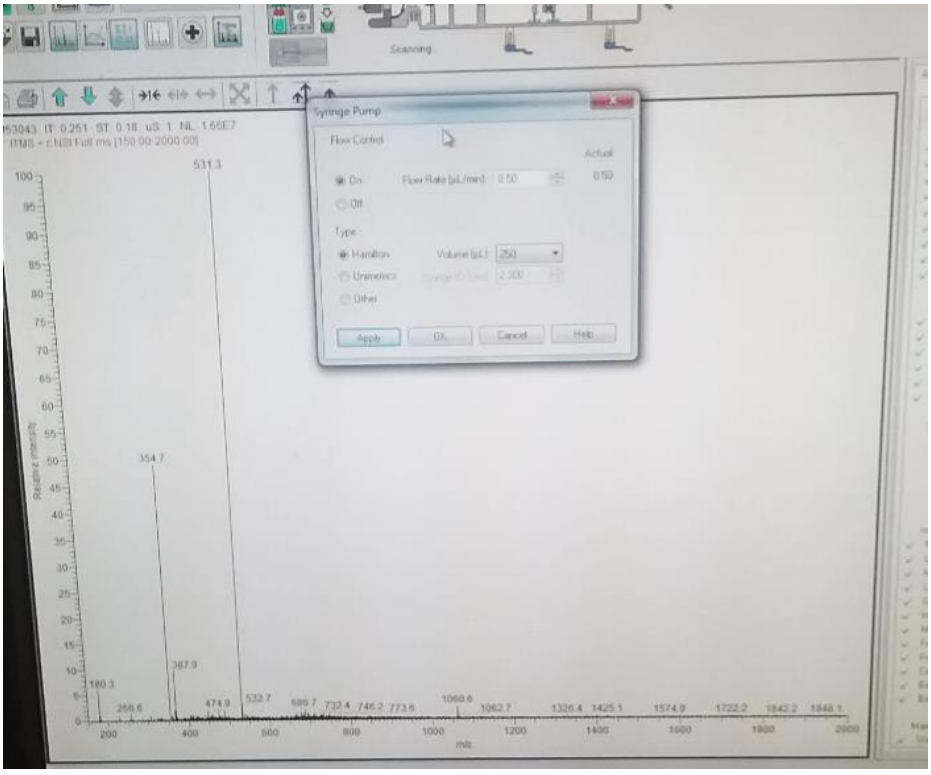

Figure 4.21 Pulled Emitter 2, Bradykinin Spectrum

The above image was captured while attempting to optimize emitter 2. Signal intensity was high (10^6-10^7). This is especially notable for an in-house pulled tip.

Two standards were purchased from Sigma-Aldrich to monitor ESI and NSI performance: an MSQC1 MS Qual/Quant QC Mix LC-MS Calibration Standard and an MSQC2 QCAL Peptide Mix. The LC-CAL contained six trypsin-digested human proteins and a selection of corresponding stable isotope labeled peptides to be able to monitor both qualitative and quantitative operation.

Table 4.4 LC-CAL Contents (*54*)

Protein[1] (UniProt Accession Number)	Calc'd MW (Da)	Approximate protein per vial[2]		Corresponding SIL peptide sequences[3]	SIL Peptide Content per vial (pmol)[4]	Theoretical Ratio[5] Light : Heavy (Protein : SIL Peptide)
		pmol	µg			
Carbonic Anhydrase I (P00915)	28739	100	2.9	GGPFSDSY[R]	100	1
				VLDALQAI[K]	50	2
Carbonic Anhydrase II (P00918)	29115	100	2.9	AVQQPDGLAVLGIFL[K]	10	10
				SADFTNFDP[R]	2	50
NAD(P)H dehydrogenase (P15559)	30736	20	0.62	EGHLSPDIVAEQ[K]	20	1
				ALIVLAHSE[R]	10	2
C-reactive Protein (P02741)	23047	20	0.46	ESDTSYVSL[K]	2	10
				GYSIFSYAT[K]	0.4	50
Peptidyl-prolyl cis-trans isomerase A (P62937)	20176	4	0.08	FEDENFIL[K]	8	0.5
				VSFELFAD[K]	4	1
				TAENF[R]	2	2
Catalase (P04040)	59625	4	0.24	GAGAFGYFEVTHDIT[K]	20	0.2
				FSTVAGESGSADTV[R]	0.4	10
				NLSVEDAA[R]	0.08	50

The LC-CAL 7-8 µg of dried peptide was reconstituted as suggested by the manufacturer. 20 µL of 20% ACN, 0.1% formic acid solution was added to the vial making a concentration of about 0.375 µg/µL. It was mixed by vortexing. 10 µL of this reconstitution was diluted in an additional 490 µL of 20% ACN, 0.1% formic acid, so 10 µL of this diluted solution contains about 1% of the total dried peptide (0.075 µg). 1% of the vial contents was recommended for first analysis; the injection volume was 10 µL for runs of this standard.

The peptide mix contained 25 µg of lyophilized peptides of a trypsin digestion of a recombinantly expressed (*E. coli*) concatamer QCAL protein.

Table 4.5 Peptide Mix peptides and Molecular Masses (*55*)

Peptide #	Sequence	Calibration Element	$[M+H]^{+}$	$[M+2H]^{2+}$	$[M+3H]^{3+}$
1	GLVK	m/z	416.2872	208.6475	139.4343
2	FVVPR	m/z, LC	617.3775	309.1926	206.4644
3	ALELFR	m/z, LC	748.4357	374.7217	250.1504
4	IGDYAGIK	m/z, LC	836.4517	418.7298	279.4891
5	EALDFFAR	m/z, LC	968.4841	484.7459	323.4999
6 (*)	ALVALVLVPR	Expression System	1050.7039	525.8558	350.9065
7	YLGYLEQLLR	m/z, LC	1267.7050	634.3564	423.2402
8	AVMDDFAAFVEK	m/z, LC, Met Ox	1342.6352	671.8215	448.2169
9 (*)	ALVALVHHHHHH	Expression System	1407.751	704.3794	469.9222
10	GVNDNEEGFFSAK	guanidation	1413.6286	707.3182	471.8814
11	GVNDNEEGFFSAR	Linearity of response	1441.6347	721.3213	481.2168
12	VLYPNDNFFEGK	m/z, LC	1442.6955	721.8517	481.5704
13	AVMMDDFAAFVEK	m/z, LC, Met Ox	1473.6757	737.3418	491.8971
14	GGVNDNEEGFFSAR	Linearity of response	1498.6562	749.8320	500.2239
15 (*)	GSLEVLFQGPIEGR	Expression System	1501.8014	751.4046	501.2723
16	GGGVNDNEEGFFSAR	Linearity of response	1555.6777	778.3427	519.2311
17 (*)	TENLYFQGDDDDK	Expression System	1559.6501	780.3289	520.5552
18	AVMMMDDFAAFVEK	m/z, LC, Met Ox	1604.7162	802.8620	535.5772
19	LFTFHADI(Cam)TLPDTEK (Cam: alkylation with iodoacetamide)	m/z, LC	1907.9213	954.4645	636.6456
20	VFDEFQPLVEEPQNLIR	m/z, resolution, LC	2073.0656	1037.0367	691.6937
21	VFDEFKPLVEEPQNLIR	m/z, resolution, LC	2073.1020	1037.0549	691.7059
22	VFDEFKPLVKPEEPQNLIR	m/z, resolution, LC	2298.2497	1149.6288	766.7551
23	VFDEFKPLVKPEEKPQNLIR	m/z, resolution, LC	2426.3447	1213.6762	809.4534
24	VFDEFKPLVKPEEKPQNKPLIR	m/z, resolution, LC	2651.4924	1326.2501	884.5027
25	VFKPDEFKPLVKPEEKPQNKPLIR	m/z, resolution, LC	2876.6401	1438.8240	959.5519
26 (*)	TENLYFQGDDDDKALVALVHHHHHH	Expression System	2948.3827	1474.6953	983.4661
27	VFKPDEFKPLVKPEEKPQNKPLIKPR	m/z, resolution, LC	3101.7878	1551.3978	1034.6011

(*) Peptides 6, 9, 15, 17, and 26 are derived from the poly-histidine tag which is used for affinity purification of the protein, and from the linker sequence. These peptides are not part of the intended product performance features of QCAL. However, because these peptides do give peaks in LC analysis, they are listed here for clarity.

The peptide mix was used reconstituted per the manufacturer's instructions. 22.7 µL of 20% ACN, 0.1% formic acid solution was added to the contents of a vial and the vial was vortexed to mix. An additional 227 µL of ACN/formic acid solution was added, and the vial was vortexed again. 10 uL was the injection volume for this standard during LC-MS runs.

Right after the flow splitter was removed, and the NSI source was replaced by ESI, a 100-minute direct gradient run of the peptide mix was performed. Since the flow split was removed, the flow rate was 400 µL/min throughout the run, but the gradient was the same as it was when NSI was employed. The method details can be viewed in sections 2.2 and 2.3. SEQUEST matched 98.92% of the QCAL protein sequence and gave it a score of 71.84. The number of spectral

matches for each peptide was much lower than for the successful NSI runs of the *E. coli* shotgun and purified ArsR. The chromatogram can be viewed below.

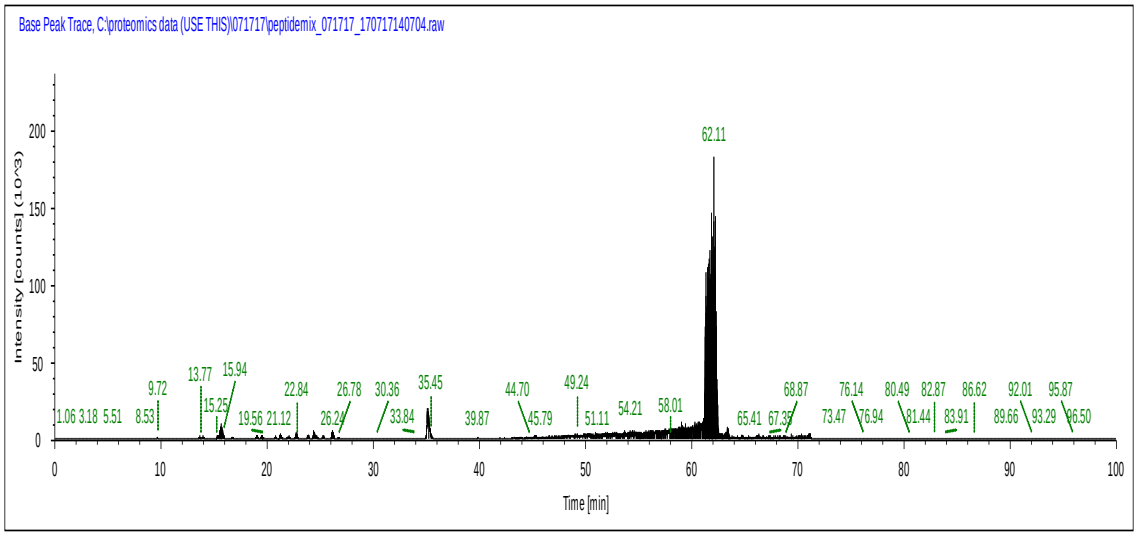

Figure 4.22 Chromatogram of Peptide Mix, 100-Minute Direct Method, ESI, 5.0 kV

No high confidence peptide sequences were identified as eluting during the largely visible peak around 62 minutes. The intensity counts for identified sequences were generally lower than 100. There was a significantly less amount of unwanted residue coming off the analytical column than before it was replaced.

Next, the peptide mix and LC-CAL solutions were run with the new 60-minute gradient.

Proteins | Peptides | Search Input | Result Filters | Peptide Confidence | Search Summary

	Accession	Description	Score	Coverage	# Proteins	# UniquePeptides	# Peptides	# PSMs	# AAs	MW [kDa]	calc. pI
1	QCAL	from MSQC2	234.76	100.00 %	1	27	49	644	465	52.2	4.77

	A2	Sequence	# PSMs	# Proteins	# Protein Groups	Protein Group Accessions	Modifications	ΔCn	XCorr	Charge	MH+ [Da]	ΔM [ppm]	RT [min]
1		VFDEFKPLVKPEEKPQNKPLIR	2	1	1	QCAL		0.0000	4.89	3	2652.80454	494.75	11.38
2		GGGVNDNEEGFFSAR	20	1	1	QCAL		0.0000	4.21	2	1552.76396	-1876.20	11.73
3		VFKPDEFKPLVKPEEKPQNK...	2	1	1	QCAL		0.0000	4.03	3	3102.85874	345.23	10.25
4		AVMMDDFAAFVEK	19	1	1	QCAL		0.0000	4.02	3	1474.76657	739.93	15.24
5		aVmMMDDFAAFVEK	13	1	1	QCAL	N-Term(Acetyl); M3(Oxidat...	0.0000	4.01	2	1663.95415	740.90	16.28
6		LFTFHADICTLPDTEK	9	1	1	QCAL	C9(Carbamidomethyl)	0.0000	4.00	3	1909.80997	989.16	13.33
7		AVmMMDDFAAFVEK	15	1	1	QCAL	M3(Oxidation)	0.0000	3.97	2	1621.15911	276.55	15.45
8		AVmmMDDFAAFVEK	13	1	1	QCAL	M3(Oxidation); M4(Oxidati...	0.0000	3.95	2	1637.25444	335.17	13.94
9		TENLYFQGDDDDK	9	1	1	QCAL		0.0000	3.86	2	1561.45024	1153.13	10.35
10		AVmmDDFAAFVEK	4	1	1	QCAL	M3(Oxidation); M4(Oxidati...	0.0000	3.78	2	1507.06035	925.76	12.78

Figure 4.23 Peptide Mix, SEQUEST Report Example, 60-minute direct method, ESI

Fixing the source improved signal and resulted in SEQUEST scoring the QCAL protein well. A score of 234.76 and a 100% sequence coverage was reported. This standard run had the highest scoring protein in SEQUEST when compared to other ESI runs.

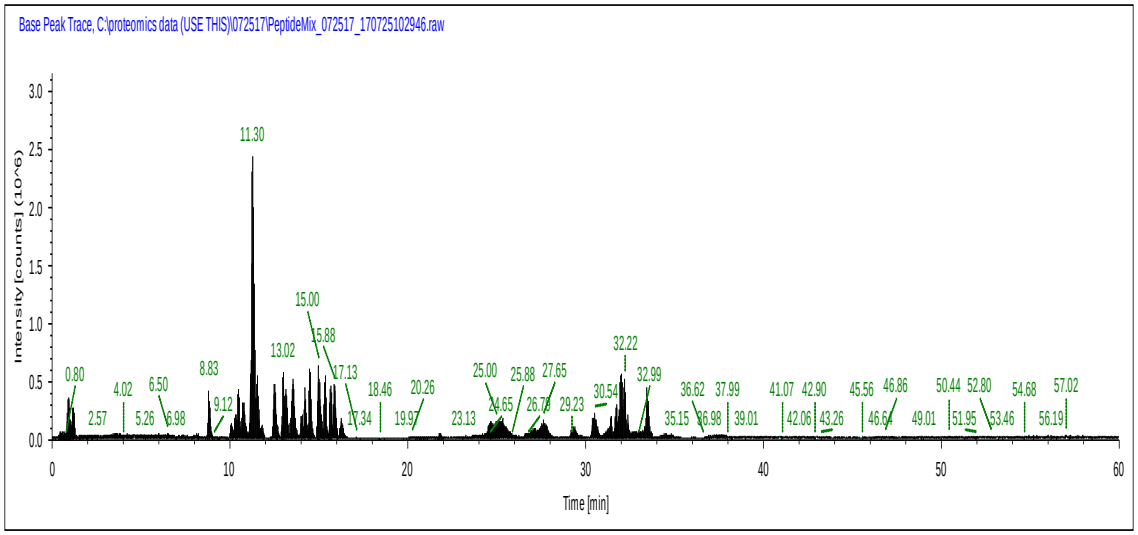

Figure 4.24 Chromatogram of Peptide Mix, 60-Minute Direct Method, ESI, 5.0 kV

Some spectra had higher ion counts than the previous run, increasing matching confidences, but there was still a problem with low signal overall. Next, the LC-CAL was run with the 60-minute direct gradient and then again with the altered 60-minute gradient and a larger injection volume (20 µL).

Proteins | Peptides | Search Input | Result Filters | Peptide Confidence | Search Summary

	Accession	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# AAs	MW [kDa]	calc. pI
1	P00915MSQC1	Carbonic anhydrase 1 (Chain 2-261) - Homo sapiens (Hu...	17.41	56.15 %	1	9	22	166	260	28.7	7.12
2	P00918MSQC1	Carbonic anhydrase 2 (Chain 2-260) - Homo sapiens (Hu...	8.04	70.27 %	1	7	29	213	259	29.1	7.40
3	P15559MSQC1	NAD(P)H dehydrogenase [quinone] 1 (Chain 2-274) - Ho...	0.00	50.18 %	1	2	35	276	273	30.7	8.88
4	P02741MSQC1	C-reactive protein (Chain 19-224) - Homo sapiens (Huma...	0.00	45.67 %	1	2	17	128	208	23.2	5.39
5	P62937MSQC1	Peptidyl-prolyl cis-trans isomerase A (Chain 1-165, N term...	0.00	73.51 %	1	0	37	324	185	20.2	8.21
6	P04040MSQC1	Catalase (Chain 2-527) - Homo sapiens (Human) - [CATA...	0.00	77.19 %	1	0	68	471	526	59.6	7.39

Figure 4.25 LC-CAL, SEQUEST Proteins Report, 60-minute direct method, ESI

The 6 proteins were identified, but low signal prevented them from scoring very well, despite their decent sequence coverage. Additionally, since the LC-CAL provided the opportunity

to analyze a complex sample with different protein/peptide abundances, the scores of zero align with the less represented proteins.

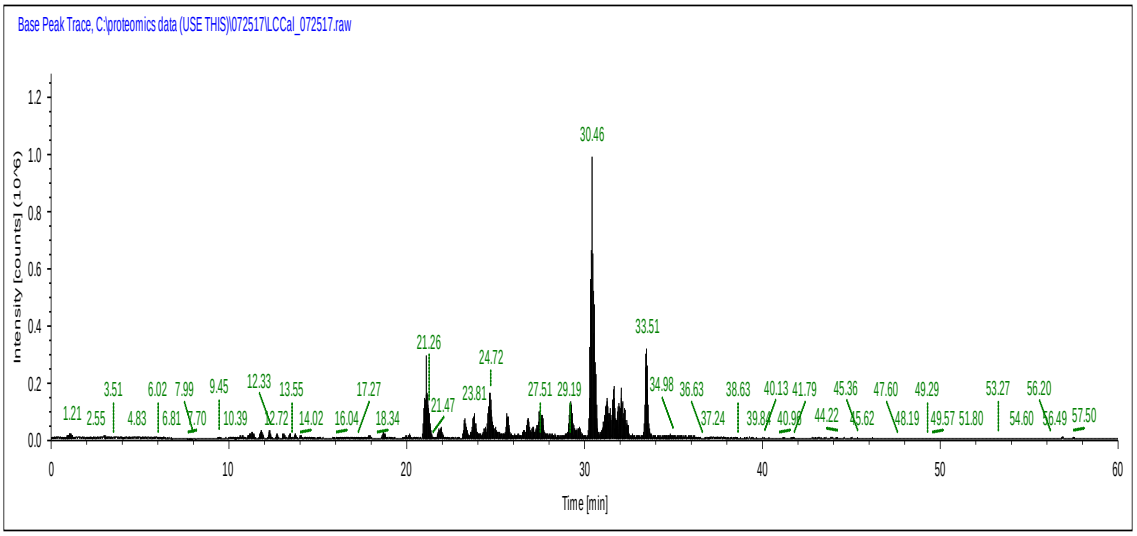

Figure 4.26 Chromatogram of LC-CAL, 60-Minute Direct Method, ESI

Low signal continued to be an issue, so the gradient was altered to extend the region of organic solvent increase.

Proteins | Peptides | Search Input | Result Filters | Peptide Confidence | Search Summary

	Accession	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# AAs	MW [kDa]	calc. pI
1	P00915MSQC1	Carbonic anhydrase 1 (Chain 2-261) - Homo sapiens (Hu...	30.96	65.77 %	1	9	22	66	260	28.7	7.12
2	P00918MSQC1	Carbonic anhydrase 2 (Chain 2-260) - Homo sapiens (Hu...	26.05	79.54 %	1	9	26	61	259	29.1	7.40
3	P04040MSQC1	Catalase (Chain 2-527) - Homo sapiens (Human) - [CATA...	3.69	73.19 %	1	5	57	137	526	59.6	7.39
4	P02741MSQC1	C-reactive protein (Chain 19-224) - Homo sapiens (Huma...	1.68	45.67 %	1	1	18	45	208	23.2	5.39
5	P15559MSQC1	NAD(P)H dehydrogenase [quinone] 1 (Chain 2-274) - Ho...	1.68	52.75 %	1	2	31	88	273	30.7	8.88
6	P62937MSQC1	Peptidyl-prolyl cis-trans isomerase A (Chain 1-165, N-term...	0.00	59.46 %	1	0	28	120	185	20.2	8.21

Figure 4.27 LC-CAL, SEQUEST Proteins Report, Altered 60-Minute Gradient, ESI 20 µL Injection

Still, all 6 proteins were identified. The scores increased, leaving only one protein of the six with a score of zero. Interestingly, one of the lowest abundant proteins in the sample scored higher than the intermediate abundance proteins. The sequence coverages shifted; they improved for some and decreased for others.

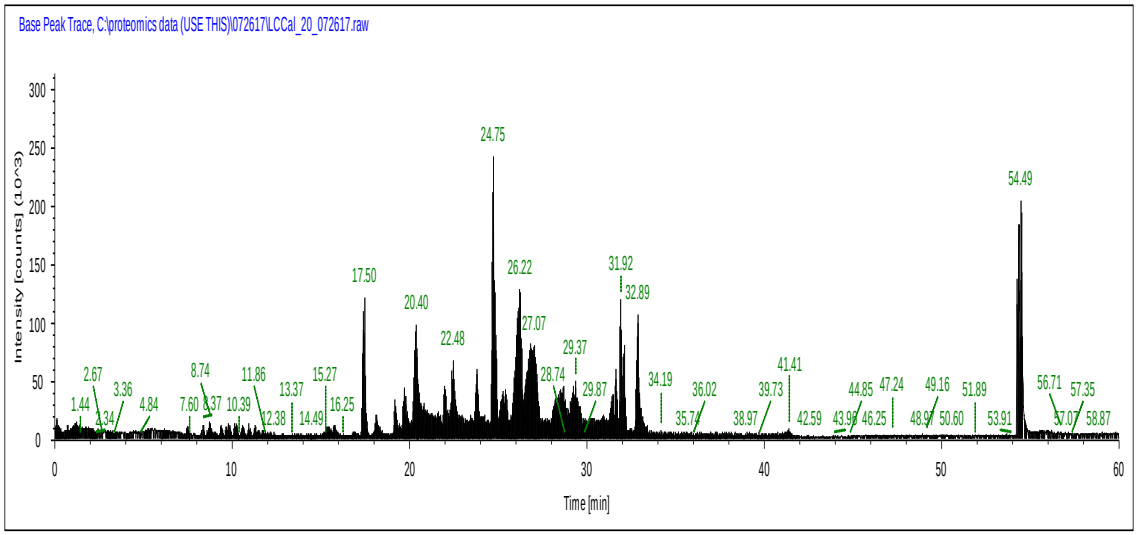

Figure 4.28 Chromatogram of LC-CAL, Altered 60-Minute Gradient, ESI, 20 µL Injection

There continued to be issues with low signal, but it is possible that the altered gradient aided in washing more hydrophobic peptides off the column in time.

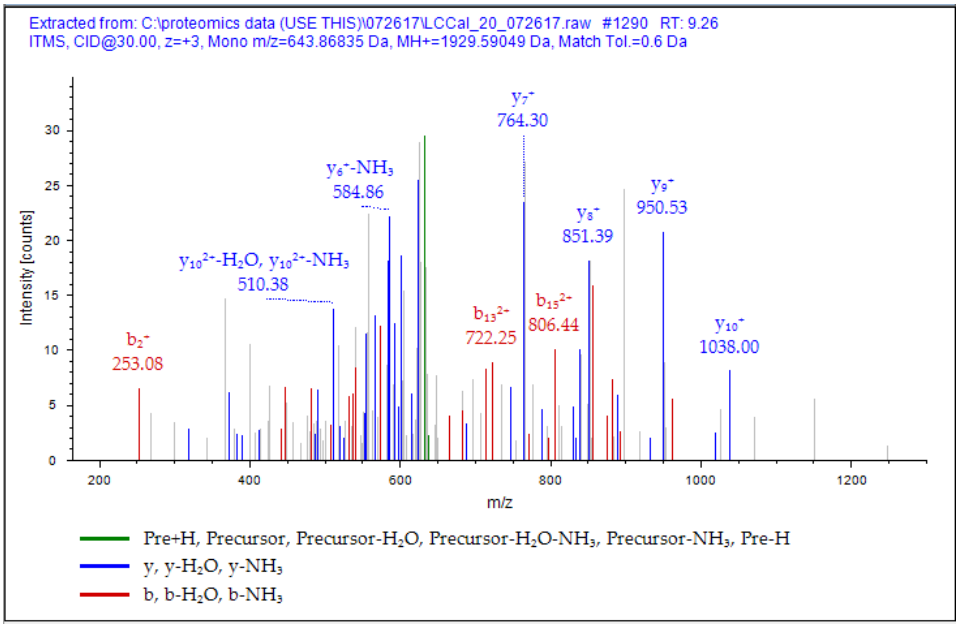

Figure 4.29 Product Ion Spectrum Matched to HDTSLKPISVSYNPATAK Annotated by SEQUEST, LC-CAL, Altered 60-Minute Gradient, ESI, 20 µL Injection

A peptide sequence with a high Xcorr, HDTSLKPISVSYNPATAK, derived from carbonic anhydrase 1, was matched to a fragmentation spectrum with ion counts below 30. There are many other peaks in this spectrum that were not matched to fragment ions.

In response to signal issues, the ESI was optimized by systematic adjustments of flow rate and source position. The 10^{-5}M bradykinin solution was used. Its doubly and triply-charged ions were used to tune. The source position was adjusted to B, yielding the best signal at 400 µL/min flow rates. The signal average for the base peak (531 m/z) fluctuated between 10^6 and 10^7. ArsR samples were run after this, revealing, through more confident results, that another enzyme needed to be used in conjunction with trypsin (Section 3.1). Another LC-CAL run was performed upon returning in September to ensure the setup was performing correctly. The 100-minute method was used.

Proteins | Peptides | Search Input | Result Filters | Peptide Confidence | Search Summary

	Accession	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# AAs	MW [kDa]	calc. pI
1	P00915MSQC1	Carbonic anhydrase 1 (Chain 2-261) - Homo sapiens (Hu...	16.41	84.23 %	1	8	33	176	260	28.7	7.12
2	P00918MSQC1	Carbonic anhydrase 2 (Chain 2-260) - Homo sapiens (Hu...	12.56	88.03 %	1	5	42	206	259	29.1	7.40
3	P02741MSQC1	C-reactive protein (Chain 19-224) - Homo sapiens (Huma...	1.61	45.67 %	1	2	20	129	208	23.2	5.39
4	P15559MSQC1	NAD(P)H dehydrogenase [quinone] 1 (Chain 2-274) - Ho...	0.00	78.39 %	1	0	49	258	273	30.7	8.88
5	P62937MSQC1	Peptidyl-prolyl cis-trans isomerase A (Chain 1-165, N term...	0.00	100.00 %	1	2	44	309	185	20.2	8.21
6	P04040MSQC1	Catalase (Chain 2-527) - Homo sapiens (Human) - [CATA...	0.00	91.44 %	1	1	95	468	526	59.6	7.39

Figure 4.30 LC-CAL, SEQUEST Proteins Report, 100-Minute Direct Method, ESI

Sequence coverages generally improved from the previous run, but the protein scores decreased. The return to the 100-minute direct gradient may have had an impact. The system was still performing as expected, but the results continually pointed out a major problem: low signal. ESI was applied to less complex sample mixtures, like the peptide mix or purified ArsR, and gave higher scores than with this protein mixture or shotgun samples.

Chapter 5: Conclusions and Future Work

NSI drastically improved ionization efficiency and detection when it was functioning correctly. ESI cannot compete with NSI's contribution to signal increases, resulting in high SEQUEST scores and confident identifications. In order to successfully perform proteomic experiments, NSI should absolutely be revisited, optimized, and routinely employed. Emitters and columns should be fabricated in-house to save money and to allow for rapid accommodation to changing experimental needs. Until ion counts increase, the results, especially for shotgun analyses, will not provide enough useful information. Additionally, the HPLC gradients currently in use could be improved to optimize separation. Gradient and method evaluations cannot be made effectively until the source is functioning properly.

With ESI, it was determined that shotgun sample preparation and analysis is not effective to analyze the low-abundance ArsR peptides. Perhaps with NSI, this could be countered, but detection would also benefit from protein separation prior to digestion. If ArsR is separated from other proteins in the cell, sample complexity is reduced. In the future, gel electrophoresis should be considered as a method to semi-purify ArsR. Successfully his-tagged *H. pylori* mutants will also aid in this endeavor because they will facilitate the use of an affinity column. NSI optimization will improve chances of detecting ArsR peptides with higher confidence. ESI runs of both purified ArsR and the shotgun sample had rather low ion counts; NSI could help improve this.

The use of a spin column for sample clean-up does not separate the majority of ArsR peptides from contaminants. Instead, a significant amount of ArsR fragments were detected in the flow-through from this procedure. ArsR is not hydrophobic enough to use this purification method. Another sample clean-up should be considered. Sample clean-up should not be neglected because of contaminants' contributions to ionization suppression.

Important considerations should also include the effects of using Asp-N as a proteolytic cleavage enzyme while trying to identify modification on aspartic acid residues. Cleaving at the residues of interest may affect the modifications that could be present. Additionally, in successfully mutated (D52E or D52N) *H. pylori*, cleavage at the 52nd residue will not be initiated by Asp-N. The effect on the length of the resulting peptides should be evaluated. The use of trypsin provides two sites for protonation, but Asp-N does not do the same. ETD needs to be utilized to analyze phosphorylation and other translational modifications because CID will remove these labile groups. ETD requires a doubly-charged precursor ion. Some of the precursor ions containing residues of interest were doubly-charged, but not all of them. This will be a key in moving forward with this project.

References

(*1*) Merriam Webster. *Proteome.* **1995**. https://www.merriam-webster.com/dictionary/proteome?utm_campaign=sd&utm_medium=serp&utm_source=jsonld (accessed August 17, 2017).

(*2*) Graves, P. R.; Haystead, T. A. Molecular Biologist's Guide to Proteomics. *Microbiol. Mol. Rev.* **2002**, 66, 39-63.

(*3*) Mitulovic, G. New HPLC Techniques for Proteomic Analysis: A Short Overview of Latest Developments. *J. Liq. Chromatogr. Relat. Technol.* **2015**, 38, 390-403.

(*4*) Allison, L. A. RNA Processing and Post-Transcriptional Gene Regulation. In *Fundamentals of Molecular Biology*, 2nd ed.; John Wiley and Sons, Inc.: USA, 2012; pp -448.

(*5*) Fosgerau, K.; Hoffmann, T. Peptide Therapeutics: Current Status and Future Direction. *Drug Discov. Today* **2015**, 20, 122-128

(*6*) Bruno, B. J.; Miller, G. D.; Lim, C. S. Basics and Recent Advances in Peptide and Protein Drug Delivery. *Ther. Deliv.* **2013**, 4, 1443-1467.

(*7*) Araúo F.; das Neves, J.; Martins, J. P.; Granja, P. L.; Santos, H. A.; Sarmento, B. Functionalized Materials for Multistage Platforms in the Oral Delivery of Biopharmaceuticals. *Prog. Mater. Sci.* **2017**, 89, 306-344.

(*8*) Aebersold, R.; Mann, M. Mass Spectrometry-Based Proteomics. *Nature* **2003**, 422, 198-207.

(*9*) Sidoli, S.; Kulej, K.; Garcia, B. A. Why Proteomics is Not the Future Genomics and the Future of Mass Spectrometry in Cell Biology. *J. Cell Biol.* **2016**, 216, 21-24.

(*10*) Osindale, N.; Aloria, K.; Omaetxebarria, M. J.; Kratchmarova, I. Targeted Mass Spectrometry: An Emerging Powerful Approach to Unblock the Bottleneck in Phosphoproteomics. *J. Chromatogr. B* **2017**, 1055-1056, 29-38.

(*11*) Han, X.; Aslanian, A. Yates III, J. R. Mass Spectrometry for Proteomics. *Curr. Opin. Chem. Biol.* **2008**, 12, 483-490.

(*12*) Overview of Cell Lysis and Protein Extraction. https://www.thermofisher.com/us/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/overview-cell-lysis-and-protein-extraction.html (Accessed Nov 16, 2017)

(*13*) Instructions: Pierce™ Mass Spec Sample Prep Kit for Cultured Cells. Thermo Scientific

(*14*) Zhang, Y.; Fonslow, B. R.; Shan, B.; Maek, M.; Yates, J. R. III. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem. Rev.* **2013**, 113, 2343-2394.

(*15*) Roepstorff, P. Mass spectrometry based proteomics, background, status and future needs. *Protein Cell* **2012**, 3, 641-647.

(*16*) Zhang, Z; Wu, Si; Stenoein, D. L.; Pasa-Tolic, L. High-Throughput Proteomics. *Annu. Rev. Anal. Chem.* **2014**, 7, 427-454.

(*17*) Gundry, R. L.; White, M. Y.; Murray, C. I.; Kane, L. A.; Fu, Q.; Stanley, B. A.; Van Eyk, J. E. Preparation of Proteins and Peptides for Mass Spectrometry Analysis in a Bottom-Up Proteomics Workflow. *Curr. Protoc. Mol. Biol.* **2009**, 88, 1-10.

(*18*) Zaikin, V.; Halket, J. *A Handbook of Derivatives for Mass Spectrometry.* IM Publications: UK, 2009; pp 300-301.

(*19*) Sidoli, S.; Lu, C.; Coradin, M.; Wang, X.; Karch, K. R.; Ruminowicz, C.; Garcia, B. A. Metabolic labeling in middle-down proteomics allows for investigation of the dynamics of the histone code. *Epigenetics and Chromatin* **2017**, 10, 34-48.

(*20*) Washburn, M. P.; Wolters, D.; Yates, J. R., III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001**, 19, 242-247.

(*21*) Zhou, F.; Sikorski, T. W.; Ficarro, S. B.; Webber, J. T.; Marto, J. A. An Online Nanoflow RP-SAX-RP LC-MS/MS Platform for Efficient and In-depth Proteome Sequence Analysis of Complex Organisms. *Anal. Chem.* **2011**, 83, 6996-7005.

(*22*) Montoyama, A.; Venable, J. D.; Ruse, C. I.; Yates, J. R., III. Automated Ultra-High-Pressure Multidimensional Protein Identification Technology (UHP-MudPIT) for Improved Peptide Identification of Proteomic Samples. *Anal. Chem.* **2006**, 78, 5109-5118.

(*23*) Kuroda, I.; Shintani, Y.; Motokawa, M.; Abe, S.; Furuno, M.; Phosphopeptideselective Column-switching RP-HPLC with a Titania Precolumn. *Anal. Sci.* **2004**, 20, 1313-1319.

(*24*) Singhal, N.; Kumar, M.; Kanaujia, P. K.; Virdi, J. S. MALDI-TOF mass spectrometry: an emerging technology for microbal identification and diagnosis. *Frontiers in Microbiology* **2015**, 6, 1-16.

(*25*) Singhal, N.; Kumar, M.; Kanaujia, P. K.; Virdi, J. S. MALDI-TOF mass spectrometry: an emerging technology for microbal identification and diagnosis. *Frontiers in Microbiology* **2015**, 6, 1-16.

(*26*) Orgorzalek Loo, R. R.; Stevenson, T. I.; Mitchell, C.; Loo, J. A.; Andrews, P. C. Mass Spectrometry of Proteins Directly from Polyacrylamide Gels. *Anal. Chem.* **1996**, 68, 1910-1917.

(*27*) Liuni, P.; Wilson, D. J.; Understanding and optimizing electrospray ionization techniques for proteomic analysis. *Expert Rev. Proteomics* **2011**, 8, 197-209.

(*28*) Wilm, M.; Mann, M. Analytical Propertied of the Nanoelectrospray Ion Source. *Anal. Chem.* **1996**, 68, 1-8.

(*29*) Juraschek, R.; Dulks, T.; Karas, M. Nanoelectrospray- More Than Just a Minimized-Flow Electrospray Ionization Source. *J. Am. Soc. Mass Spectrom.* **1999**, 10, 300-308.

(*30*) Douglas, D.J.; Frank, A. J.; Mao, D. Linear Ion Traps in Mass Spectrometry. In *Mass Spectrometry Reviews* [Online}; Wiley Periodicals, Inc, 2005; Vol. 24, pp 1-29. http://onlinelibrary.wiley.com/doi/10.1002/mas.20004/epdf (Accessed Nov 17, 2017).

(*31*) Guilhaus, M.; Principles and Instrumentation in Time-of-Flight Mass Spectrometry: Physical and Instrumental Concepts. *J. Mass Spectrom.* **1995**, 30, 1519-1532.

(*32*) Hu, Y.; Wan, J. H.; Li, X. Y.; Zhu, Y.; Graham, D. Y.; Lu, N. H. Systematic review with meta-analysis: the global reccurrence rate of *Helcobacter pylori*. *Aliment. Pharmacol. Ther.* **2017**, 46, 773-779.

(*33*) Salama, N. R.; Hartung, M. L.; Muller, A. Life in the human stomach: persistence strategies of the bacterial pathogen *Helicobacter pylori*. *Nat. Rev. Microbiol.* **2013**, 11, 385-399.

(*34*) Wroblewski, L. E.; Peek, R. M., Jr.; Wilson, K. T. *Helicobacter pylori* and Gastric Cancer: Factors That Modulate Disease Risk. *Clin. Microbiol. Rev.* **2010**, 23, 713-739.

(*35*) Thung, I.; Aramin, H.; Vavinskava, V.; Gupta, S.; Park, J. Y.; Crow, S. E.; Valasek, M. A. Review article: the global emergence of *Helicobacter pylori* antibiotic resistance. *Aliment. Pharmacol. Ther.* **2016**, 43, 514-533.

(*36*) Goodwin, A. C.; Weinberger, D. M.; Ford, C. B.; Nelson, J. C.; Snider, J. D.; Hall, J. D.; Paules, C. I.; Peek, R. M., Jr.; Forsyth, M. H. Expression of the *Helicobacter pylori* adhesin SabA is controlled via phase variation and the ArsRS signal transduction system. *Microbiology* **2008**, 154, 2231-2240.

(*37*) Instructions: Piece® C18 Spin Columns. Thermo Scientific.

(*38*) Eng, J. K.; McCormack, A. L.; Yates, J. R., III. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, 5, 976-989.

(*39*) Yates, J. R., III; Eng, J. K.; Clauser, K. R.; Burlingame, A. L. Search of Sequence Databases with Uninterpreted High-Energy Collision-Induced Dissociation Spectra of Peptides. *J. Am. Soc. Mass Spectrom.* **1996**, 7, 1089-1098.

(*40*) *Thermo Proteome Discoverer User Guide: Software Version 2.1* [Online] **2015**; Thermo Fisher Scientific Incorporated: USA; https://tools.thermofisher.com/content/sfs/manuals/Man-XCALI-97751-Proteome-Discoverer-21-User-ManXCALI97751-A-EN.pdf (accessed Nov 16, 2017)

(*41*) Zanotti, G.; Cendron, L. Structural and functional aspects of the *Helicobacter pylori* secretome. *World J. Gastroenterol.* **2014**, 20, 1402-1423.

(*42*) Smith, P. K.; Krohn, R. I.; Hermanson, G. T.; Mallia, A. K.; Gartner, F. H.; Provenzano, M. D.; Fujimoto, E. K.; Goeke, N. M.; Olsen, B. J.; Klenk, D.C. Measurement of protein using bicinchoninic acid. *Anal Biochem* **185**, 150, 76-85.

(*43*) *PicoView Nanospray Source User Manual: Model PV-550*; New Objective: USA, 2009.

(*44*) David, M. T.; Stahl, D. C.; Hefta, S. A.; Lee, T. D.; A Microscale Electrospray Interface for On-Line, Capillary Liquid Chromatography/Tandem Mass Spectrometry of Complex Peptide Mixtures. *Anal. Chem.* **1995**, 67, 4549-4556.

(*45*) Valaskovic, N. L.; Kelleher, N. L.; Little, D. P.; Aaserud, D. J.; McLafferty, F. Attomole-Sensitivity Electrospray Source for Large-Molecule Mass Spectrometry. *Anal. Chem.* **1995**, 67, 3802-3805.

(*46*) Viberg, P.; Nilsson, S.; Skog, K. Nanospray Mass Spectrometry with Indirect Conductive Graphite Coating. *Anal. Chem.* **2004**, 76, 4241-4244.

(*47*) Ek, P.; Roeraade, J. New Method for Fabrication of Fused Silica Emitters with Submicrometer Orifices for Nanoelectrospray Mass Spectrometry. *Anal. Chem.* **2011**, 83, 7771-7777

(*48*) Kelly, R. T.; Page, J. S.; Luo, Q.; Moore, R. J.; Tang, K.; Smith, R. D. Chemically Etched Open Tubular and Monolithic Emitters for Nanoelectrospray Ionization Mass Spectrometry. *Anal. Chem.* **2006**, 78, 7796-7801.

(*49*) Wetterhall, M.; Nilsson, S.; Markides, K. E.; Bergquist, J. A conductive Polymeric Material Used for Nanospray Needle and Low-Flow Sheathless Electrospray Ionization Applications. *Anal. Chem.* **2002**, 74, 239-245.

(*50*) Ahern, J. C.; Balanchandran, W. Experimental Electrohydrodynamic Nanospray Production Using Drawn Glass Capillaries. *Particulate Science and Technology* 2006, 24, 271-280.

(*51*) UWPR Advancing Proteomics: Packing Capillary Columns and Pre-columns (traps). *University of Washington Proteomics Resource*, 2013.

(*52*) Maziarz, E. P., III; Lorenz, S. A.; White, T. P.; Wood, T. D. Polyaniline: A conductive Polymer Coating for Durable Nanospray Emitters. *J. Am. Soc. Mass Spectrom.* **2000**, 11, 659-663.

(*53*) Wetterhall, M.; Klett, O.; Markides, K. E.; Nyholm, L.; Bergquist, J. A comparison of the electrochemical stabilities of metal, polymer, and graphite coated nanospray emitters. *Analyst*, **2003**, 128, 728-733.

(*54*) Product Information: MS Qual/Quant QC Mix. Sigma-Aldrich.

(*55*) Product Information: MS QCAL Peptide Mix. Sigma-Aldrich.